

ON KAC'S CHAOS AND RELATED PROBLEMS

M. HAURAY AND S. MISCHLER

ABSTRACT. This paper is devoted to establish quantitative and qualitative estimates related to the notion of chaos as firstly formulated by M. Kac [41] in his study of mean-field limit for systems of N undistinguishable particles as $N \rightarrow \infty$.

First, we quantitatively liken three usual measures of *Kac's chaos*, some involving the all N variables, other involving a finite fixed number of variables. The cornerstone of the proof is a new representation of the Monge-Kantorovich-Wasserstein (MKW) distance for symmetric N -particle probability measures in terms of the distance between the law of the associated empirical measures on the one hand, and a new estimate on some MKW distance on probability measures spaces endowed with a suitable Hilbert norm taking advantage of the associated good algebraic structure.

Next, we define the notion of *entropy chaos* and *Fisher information chaos* in a similar way as defined by Carlen et al [17]. We show that *Fisher information chaos* is stronger than *entropy chaos*, which in turn is stronger than *Kac's chaos*. More importantly, with the help of the HWI inequality of Otto-Villani, we establish a quantitative estimate between these quantities, which in particular asserts that *Kac's chaos* plus *Fisher information bound* implies *entropy chaos*.

We then extend the above quantitative and qualitative results about chaos in the framework of probability measures with support on the *Kac's spheres*, revisiting [17] and giving a possible answer to [17, Open problem 11]. Additionally to the above mentioned tool, we use and prove an optimal rate local CLT in L^∞ norm for distributions with finite 6-th moment and finite L^p norm, for some $p > 1$.

Last, we investigate how our techniques can be used without assuming chaos, in the context of probability measures mixtures introduced by De Finetti, Hewitt and Savage. In particular, we define the (level 3) Fisher information for mixtures and prove that it is l.s.c. and affine, as that was done in [64] for the level 3 Boltzmann's entropy.

April 2, 2013

Keywords: Kac's chaos, Monge-Kantorovich-Wasserstein distance, Entropy chaos, Fisher information chaos, CLT with optimal rate, probability measures mixtures, De Finetti, Hewitt and Savage theorem, Mean-field limit, quantitative chaos, qualitative chaos, entropy.

AMS Subject Classification:

26D15 Inequalities for sums, series and integrals,

94A17 Measures of information, entropy,

60F05 Central limit and other weak theorems,

82C40 Kinetic theory of gases.

43A15 L^p -spaces and other function spaces on groups, semigroups

52A40 Inequalities and extremum problems

CONTENTS

1. Introduction and main results	2
----------------------------------	---

2. Kac's chaos	10
3. Entropy chaos and Fisher information chaos	28
4. Probability measures on the "Kac's spheres"	37
5. On mixtures according to De Finetti, Hewitt and Savage	55
References	77

1. INTRODUCTION AND MAIN RESULTS

The Kac's notion of chaos rigorously formalizes the intuitive idea for a family of stochastic valued vectors with N coordinates to have asymptotically independent coordinates as N goes to infinity. We refer to [67] for an introduction to that topics from a probabilistic point of view, as well as to [54] for a recent and short survey.

Definition 1.1. [41, section 3] *Consider $E \subset \mathbb{R}^d$, $f \in \mathbf{P}(E)$ a probability measure on E and $G^N \in \mathbf{P}_{\text{sym}}(E^N)$ a sequence of probability measures on E^N , $N \geq 1$, which are invariant under coordinates permutations. We say that (G^N) is f -Kac's chaotic (or has the "Boltzmann property") if*

$$(1.1) \quad \forall j \geq 1, \quad G_j^N \rightharpoonup f^{\otimes j} \quad \text{weakly in } \mathbf{P}(E^j) \quad \text{as } N \rightarrow \infty,$$

where G_j^N stands for the j -th marginal of G^N defined by

$$G_j^N := \int_{E^{N-j}} G^N dx_{j+1} \dots dx_N.$$

Interacting N -indistinguishable particle systems are naturally described by exchangeable random variables (which corresponds to the fact that their associated probability laws are symmetric, i.e. invariant under coordinates permutations) but they are not described by random variables with independent coordinates (which corresponds to the fact that their associated probability laws are tensor products) except for situations with no interaction! Kac's chaos is therefore a well adapted concept to formulate and investigate the infinite number of particles limit $N \rightarrow \infty$ for these systems as it has been illustrated by many works since the seminal article by Kac [41]. Using the above definition of chaos, it is shown in [41, 49, 50, 35, 55] that if $f(t)$ evolves according to the nonlinear space homogeneous Boltzmann equation, $G^N(t)$ evolves according to the linear Master/Kolmogorov equation associated to the stochastic Kac-Boltzmann jumps (collisions) process and $G^N(0)$ is $f(0)$ -chaotic, then for any later time $t > 0$ the sequence $G^N(t)$ is also $f(t)$ -chaotic: in other words *propagation of chaos* holds for that model. As it is explained in the latest reference and using the uniqueness of statistical solutions proved in [2], some of these propagation of chaos results can be seen as an illustration of the "BBGKY hierarchy method" whose most famous success is the Lanford's proof of the "Boltzmann-Grad limit" [43].

In order to investigate quantitative version of Kac's chaos, the above weak convergence in (1.1) can be formulated in terms of the Monge-Kantorovich-Wasserstein (MKW) transportation distance between G_j^N and $f^{\otimes j}$. More precisely, given d_E a bounded distance on E , we define the normalized distance d_{E^j} on E^j , $j \in \mathbb{N}^*$, by setting

$$(1.2) \quad \forall X = (x_1, \dots, x_j), Y = (y_1, \dots, y_j) \in E^j \quad d_{E^j}(X, Y) := \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i),$$

and then we define W_1 (without specifying the dependence on j) the associated MKW distance in $\mathbf{P}(E^j)$ (see the definition (2.2) below). With the notations of Definition 1.1, G^N is f -Kac's chaotic if, and only if,

$$\forall j \geq 1, \quad \Omega_j(G^N; f) := W_1(G_j^N, f^{\otimes j}) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Let us introduce now another formulation of Kac's chaos which we firstly formulate in a probabilistic language. For any $X = (x_1, \dots, x_N) \in E^N$, we define the associated empirical measure

$$(1.3) \quad \mu_X^N(dy) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dy) \in \mathbf{P}(E).$$

We say that an exchangeable E^N -valued random vector \mathcal{X}^N is f -chaotic if the associated $\mathbf{P}(E)$ -valued random variable $\mu_{\mathcal{X}^N}^N$ converges to the deterministic random variable f in law in $\mathbf{P}(E)$:

$$(1.4) \quad \mu_{\mathcal{X}^N}^N \Rightarrow f \quad \text{in law as } N \rightarrow \infty.$$

In the framework of Definition 1.1, the convergence (1.4) can be equivalently formulated in the following way. Introducing $G^N := \mathcal{L}(\mathcal{X}^N)$ the law of \mathcal{X}^N , the exchangeability hypothesis means that $G^N \in \mathbf{P}_{sym}(E^N)$. Next the law $\hat{G}^N := \mathcal{L}(\mu_{\mathcal{X}^N}^N)$ of $\mu_{\mathcal{X}^N}^N$ is nothing but the (unique) measure $\hat{G}^N \in \mathbf{P}(\mathbf{P}(E))$ such that

$$\langle \hat{G}^N, \Phi \rangle = \int_{E^N} \Phi(\mu_X^N) G^N(dX) \quad \forall \Phi \in C_b(\mathbf{P}(E)),$$

or equivalently the push-forward of G^N by the “empirical distribution” application.

Then the convergence (1.4) just means that

$$(1.5) \quad \hat{G}^N \rightarrow \delta_f \quad \text{weakly in } \mathbf{P}(\mathbf{P}(E)) \text{ as } N \rightarrow \infty,$$

where this definition does not refer anymore to the random variables \mathcal{X}^N or $\mu_{\mathcal{X}^N}^N$. It is well known (see for instance [36, section 4], [41, 69, 66] and [67, Proposition 2.2]) that for a sequence (G^N) of $\mathbf{P}_{sym}(E^N)$ and a probability measure $f \in \mathbf{P}(E)$ the three following assertions are equivalent:

- (i) convergence (1.1) holds for any $j \geq 1$;
- (ii) convergence (1.1) holds for some $j \geq 2$;
- (iii) convergence (1.5) holds;

so that in particular (1.1) and (1.5) are indeed equivalent formulations of Kac's chaos. The chaos formulation (ii) has been used since [41], while the chaos formulation (iii) is widely used in the works by Sznitman [65], see also [66, 52, 59], where the chaos property is established by proving that the “empirical process” $\mu_{\mathcal{X}^N}^N$ converges to a limit process with values in $\mathbf{P}(E)$ which is a solution to a nonlinear martingale problem associated to the mean-field limit equation. Formulation (1.5) is also well adapted for proving quantitative propagation of chaos for deterministic dynamics associated to the Vlasov equation with regular interaction force [27] as well as singular interaction force [39, 38]. Let us briefly explain this point now, see also [54, section 1.1]. On the one hand, introducing the MKW transport distance $\mathcal{W}_1 := \mathcal{W}_{W_1}$ on $\mathbf{P}(\mathbf{P}(E))$ based on the MKW distance W_1 on $\mathbf{P}(E)$, (see definition (2.6) below), the weak convergence (1.5) is nothing but the fact that

$$\Omega_\infty(G^N; f) := \mathcal{W}_1(\hat{G}^N, \delta_f) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

On the other hand, for the Vlasov equation with smooth and bounded force term, it is proved in [27] that

$$(1.6) \quad \forall T > 0, \forall t \in [0, T] \quad W_1(\mu_{\mathcal{X}_t^N}^N, f_t) \leq C_T W_1(\mu_{\mathcal{X}_0^N}^N, f_0),$$

where $f_t \in \mathbf{P}(E)$ is the solution to the Vlasov equation with initial datum f_0 and $\mathcal{X}_t^N \in E^N$ is the solution to the associated system of ODEs with initial datum \mathcal{X}_0^N . Inequality (1.6) is a consequence of the fact that $t \mapsto \mu_{\mathcal{X}_t^N}^N$ solves the Vlasov equation and that a local W_1 stability result holds for such an equation. When \mathcal{X}_0 is distributed according to an initial density $G_0^N \in \mathbf{P}_{sym}(E^N)$ we may show that \mathcal{X}_t is distributed according to $G_t^N \in \mathbf{P}_{sym}(E^N)$ obtained as the transported measure along the flow associated to the above mentioned system of ODEs or equivalently G_t^N is the solution to the associated Liouville equation with initial condition G_0^N . Taking the expectation in both sides of (1.6), we get

$$\begin{aligned} \int_{E^N} W_1(\mu_Y^N, f_t) G_t^N(dY) &= \mathbb{E}[W_1(\mu_{\mathcal{X}_t^N}^N, f_t)] \\ &\leq C_T \mathbb{E}[W_1(\mu_{\mathcal{X}_0^N}^N, f_0)] = C_T \int_{E^N} W_1(\mu_Y^N, f_0) G_0^N(dY), \end{aligned}$$

for any $t \in [0, T]$. We conclude with the following quantitative chaos propagation estimate

$$\forall t \in [0, T] \quad \Omega_\infty(G_t^N; f_t) \leq C_T \Omega_\infty(G_0^N; f_0).$$

It is worth mentioning that partially inspired from [36], it is shown in [57, 55] a similar inequality as above for more general models including drift, diffusion and collisional interactions where however the estimate may mix several chaos quantification quantities as Ω_∞ and Ω_2 for instance.

There exists at least one more way to guaranty chaoticity which is very popular because that chaos formulation naturally appears in the probabilistic coupling technique, see [67], as well as [47, 12, 11] and the references therein.

Thanks to the coupling techniques we typically may show that an exchangeable E^N -valued random vector \mathcal{X}^N satisfies

$$\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N |\mathcal{X}_i^N - \mathcal{Y}_i^N|\right) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

for some E^N -valued random vector \mathcal{Y}^N with independent coordinates. Denoting by $G^N \in \mathbf{P}_{sym}(E^N)$ the law of \mathcal{X}^N , f the law of one coordinate \mathcal{Y}_i^N , and W_1 the MKW transport distance on $\mathbf{P}(E^N)$ based on the normalized distance d_{E^N} in E^N defined by (1.2), the above convergence readily implies

$$(1.7) \quad \Omega_N(G^N; f) := W_1(G^N, f^{\otimes N}) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

which in turn guaranties that (G^N) is f -chaotic. It is generally agreed that the convergence (1.7) is a strong version of chaos, maybe because it involves the all N variables, while the Kac's original definition only involves a finite fixed number of variables.

Summary of Section 2. The first natural question we consider is about the equivalence between these definitions of chaos, and more precisely the possibility to liken them in a quantitative way. The following result gives a positive answer, we also refer to Theorem 2.4 in section 2 for a more accurate statement.

Theorem 1.2 (Equivalence of measure for Kac's chaos). *For any moment order $k > 0$ and any positive exponent $\gamma < (d + 1 + d/k)^{-1}$, there exists a constant $C = C(d, k, \gamma) \in (0, \infty)$*

such that for any $f \in \mathbf{P}(E)$, any $G^N \in \mathbf{P}_{sym}(E^N)$, $N \geq 1$, and any $j, \ell \in \{1, \dots, N\} \cup \{\infty\}$, $\ell \neq 1$, there holds

$$\Omega_j(G^N; f) \leq C \mathcal{M}_k^{1/k} \left(\Omega_\ell(G^N; f) + \frac{1}{N} \right)^\gamma,$$

where $\mathcal{M}_k = M_k(f) + M_k(G_1^N)$ is the sum of the moments of order k of f and G_1^N .

It is worth emphasizing that the above inequality is definitively false in general for $\ell = 1$. The first outcome of our theorem is that it shows that, regardless of the rate, the propagation of chaos results obtained by the coupling method is of the same nature as the propagation of chaos result obtained by the “BBGKY hierarchy method” and the “empirical measures method”.

The proof of Theorem 2.4 (from which Theorem 1.2 follows) will be presented in section 2. Let us briefly explain the strategy. First, the fact that we may control Ω_j by Ω_ℓ for $1 \leq j \leq \ell \leq N$ is classical and quite easy. Next, we will establish an estimate of Ω_∞ by Ω_2 following an idea introduced in [55]: we begin to prove a similar estimate where we replace Ω_∞ by the MKW distance in $\mathbf{P}(\mathbf{P}(E))$ associated to the $H^{-s}(\mathbb{R}^d)$ norm, $s > (d+1)/2$, on $\mathbf{P}(E)$ in order to take advantage of the good algebraic structure of that Hilbert norm and then we come back to Ω_∞ thanks to the “uniform topological equivalence” of metrics in $\mathbf{P}(E)$ and the Hölder inequality. Finally, and that is the other key new result, we compare Ω_∞ and Ω_N : that is direct consequence of the following identity

$$\forall F^N, G^N \in \mathbf{P}_{sym}(E^N) \quad W_1(G^N, F^N) = \mathcal{W}_1(\hat{G}^N, \hat{F}^N)$$

applied to $F^N := f^{\otimes N}$ and a functional version of the law of large numbers.

Summary of section 3. A somewhat stronger notion of chaos can be formulated in terms of entropy functionals. Such a notion has been explicitly introduced by Carlen, Carvahlo, Le Roux, Loss, Villani in [17] (in the context of probability measures with support on the “Kac’s spheres”) but it is reminiscent of the works [42, 6]. We also refer to [64, 53, 13, 14] where the N particles entropy functional below is widely used in order to identify the possible limits for a system of N particles as $N \rightarrow \infty$. Consider $E \subset \mathbb{R}^d$ an open set or the adherence of a open space, in order that the gradient of a function may be well defined. For a (smooth and/or decaying enough) probability measure $G^N \in \mathbf{P}_{sym}(E^N)$ we define (see section 3 for the suitable definitions) the Boltzmann’s entropy and the Fisher information by

$$H(G^N) := \frac{1}{N} \int_{E^N} G^N \log G^N dX, \quad I(G^N) := \frac{1}{N} \int_{E^N} \frac{|\nabla G^N|^2}{G^N} dX.$$

It is worth emphasizing that contrarily to the most usual convention, adopted for instance in [17, Definition 8], we have put the normalized factor $1/N$ in the definitions of the entropy and the Fisher information. Moreover we use the same notation for these functionals whatever is the dimension. As a consequence, we have $H(f^{\otimes N}) = H(f)$ and $I(f^{\otimes N}) = I(f)$ for any probability measures $f \in \mathbf{P}(E)$.

Definition 1.3. Consider (G^N) a sequence of $\mathbf{P}_{sym}(E^N)$ such that for $k > 0$ the k -th moment $M_k(G_1^N)$ is uniformly bounded in N , and $f \in \mathbf{P}(E)$. We say that

(a) (G^N) is f -entropy chaotic (or f -chaotic in the sense of the Boltzmann’s entropy) if

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad H(G^N) \rightarrow H(f), \quad H(f) < \infty;$$

(b) (G^N) is f -Fisher information chaotic (or f -chaotic in the sense of the Fisher information) if

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad I(G^N) \rightarrow I(f), \quad I(f) < \infty.$$

Our second main result is the following qualitative comparison of the three above notions of chaos convergence.

Theorem 1.4. *Assume $E = \mathbb{R}^d$, $d \geq 1$, or E is a bi-Lipschitz volume preserving deformation of a convex set of \mathbb{R}^d , $d \geq 1$. Consider (G^N) a sequence of $\mathbf{P}_{\text{sym}}(E^N)$ such that the k -th moment $M_k(G_1^N)$ is bounded, $k > 2$, and $f \in \mathbf{P}(E)$.*

In the list of assertions below, each one implies the assertion which follows:

- (i) (G^N) is f -Fisher information chaotic;
- (ii) (G^N) is f -Kac's chaotic and $I(G^N)$ is bounded;
- (iii) (G^N) is f -entropy chaotic;
- (iv) (G^N) is f -Kac's chaotic.

More precisely, the following quantitative estimate of the implication (ii) \Rightarrow (iii) holds:

$$(1.8) \quad |H(G^N) - H(f)| \leq C_E K \Omega_N(G^N; f)^\gamma,$$

with $\gamma := 1/2 - 1/k$, $K := \sup_N I(G^N)^{1/2} \sup_N M_k(G_1^N)^{1/k}$ and C_E is a constant depending on the set E (one can choose $C_E = 8$ when $E = \mathbb{R}^d$).

The implication (ii) \Rightarrow (iii) is the most interesting part and hardest step in the proof of Theorem 1.4. It is based on estimate (1.8) which is a mere consequence of the HWI inequality of Otto and Villani proved in [61] when $E = \mathbb{R}^d$

together with our equivalence of chaos convergences previously established. It is also the most restrictive one in term of moment bound: the implication (ii) \Rightarrow (iii) requires a k -th moment bound of order $k > 2$ while the other implications only require k -th moment bound of order $k > 0$ or no moment bound condition (we refer to the proof of Theorem 1.4 in section 3 for details). The proofs of the implications (i) \Rightarrow (ii) and (iii) \Rightarrow (iv) use the fact that the subadditivity inequalities of the Fisher information and of the entropy are saturated if and only if the probability measure is a tensor product. For functionals involving the entropy, similar ideas are classical and they have been used in [53, 37, 76, 13, 58] for instance.

We believe that this result gives a better understanding of the different notions of chaos. Other but related notions of entropy chaos are introduced and discussed in [17, 56]. The entropy chaos definition in [17], which consists in asking for point (iii) and (iv) above, is in fact equivalent to ours thanks to Theorem 1.4.

It is worth emphasizing that Theorem 1.4 may be very useful in order to obtain entropic propagation of chaos (possibly with rate estimate) in contexts where some bound on the Fisher information is available and propagation of Kac's chaos is already proved. Unfortunately, a bound on the Fisher information is not easy to propagate for N particle systems. However, for the so-called “*Maxwell molecules cross-section*”, following the proof of the fact that the Fisher information decreases along time for solutions to the homogeneous nonlinear Boltzmann equation [48, 70, 74] and for solutions to the homogeneous nonlinear Landau equation [75], it has been established that the N particle Fisher information also decreases along time for the law of solutions to the stochastic Kac-Boltzmann jumps process in [55, Lemma 7.4] and for the law of solutions to the stochastic Kac-Landau diffusion process in [20]. In these particular cases, Theorem 1.4 provides a quantitative version of the entropic propagation of chaos proved in [55], and we refer to [19, 20] for details.

Summary of Section 4.

Here we consider the framework of probability measures with support on the “Kac’s spheres” \mathcal{KS}_N defined by

$$\mathcal{KS}_N := \{V = (v_1, \dots, v_N) \in \mathbb{R}^N, v_1^2 + \dots + v_N^2 = N\},$$

as firstly introduced by Kac in [41]. Our aim is mainly to revisit the recent work [17] and to develop “quantitative” versions of the chaos analysis.

We start proving a quantified “Poincaré Lemma” establishing that the sequence of uniform probability measures σ^N on \mathcal{KS}_N is γ -Kac’s chaotic, with γ the standard gaussian on \mathbb{R} , i.e. $\gamma(v) = (2\pi)^{-1/2} \exp(-|v|^2/2)$, in the sense that we prove a rate of converge to 0 for the quantification of chaos $\Omega_N(\sigma^N; \gamma)$. We also prove that for a large class of probability densities $f \in \mathbf{P}(E)$ the corresponding sequence (F^N) of “conditioned to the Kac’s spheres product measures” (see section 4.2 for the precise definition) is f -Kac’s chaotic in the sense that we prove a rate of converge to 0 for the quantification of chaos $\Omega_2(F^N; f)$. That last result generalizes the “Poincaré Lemma” since $f = \gamma$ implies $F^N = \sigma^N$. The main argument in the last result is a (maybe new) L^∞ optimal rate version of the Berry-Esseen theorem, also called local central limit theorem, which is nothing but an accurate (but less general) version of [17, Theorem 27]. Together with Theorem 1.2, or the more accurate version of it stated in section 2, we obtain the following estimates.

Theorem 1.5. *The sequence (σ^N) of uniform probability measures on the “Kac’s spheres” is γ -Kac’s chaotic, and more precisely*

$$(1.9) \quad \forall N \geq 1 \quad \Omega_2(\sigma^N; \gamma) \leq \frac{C_1}{N}, \quad \Omega_N(\sigma^N; \gamma) \leq \frac{C_2}{N^{\frac{1}{2}}}, \quad \Omega_\infty(\sigma^N; \gamma) \leq C_3 \frac{(\ln N)^{\frac{1}{2}}}{N^{\frac{1}{2}}},$$

for some numerical constants C_i , $i = 1, 2, 3$.

More generally, consider $f \in \mathbf{P}(\mathbb{R})$ with bounded moment $M_k(f)$ of order $k \geq 6$ and bounded Lebesgue norm $\|f\|_{L^p}$ of exponent $p > 1$. Then, the sequence (F^N) of associated “conditioned (to the Kac’s spheres) product measures” is f -Kac’s chaotic, and more precisely

$$(1.10) \quad \forall N \geq 1 \quad \Omega_2(F^N; f) \leq \frac{C_4}{N^{\frac{1}{2}}}, \quad \Omega_N(F^N; f) \leq \frac{C_5}{N^{\frac{1}{2}}}, \quad \Omega_\infty(F^N; f) \leq \frac{C_6}{N^{\frac{1}{2}}},$$

for any $\gamma \in (0, (2 + 2/k)^{-1})$ and for some constants $C_i = C_i(f, \gamma, k)$, $i = 4, 5, 6$.

Let us briefly discuss that last result. The question of establishing the convergence for the empirical law of large numbers associated to i.i.d. samples is an important question in theoretical statistics known as Glivenko-Cantelli theorem, and the historical references seems to be [33, 15, 71]. Next the question of establishing rates of convergence in MKW distance in the above convergence has been addressed for instance in [28, 1, 26, 62, 55, 10], while the optimality of that rates have been considered for instance in [1, 68, 26, 4]. We refer to [4, 10] and the references therein for a recent discussion on that topics. With our notations, the question consists in establishing the estimate

$$(1.11) \quad \mathbb{E}(W_1(\mu_{\mathcal{X}^N}^N, f)) = \Omega_\infty(f^{\otimes N}; f) \leq \frac{C}{N^\zeta},$$

for some constants $C = C(f)$ and $\zeta = \zeta(f)$. In the above left hand side term, \mathcal{X}^N is a E^N -valued random vector with independent coordinates with identical law f or equivalently $\mathcal{X}^N = X$ is the identity vector in E^N and \mathbb{E} is the expectation associated to the tensor product probability measure $f^{\otimes N}$. When $E = \mathbb{R}^d$, estimate (1.11) has been proved to hold with $\zeta = 1/d$, if $d \geq 3$ and $\text{supp } f$ is compact in [26], with $\zeta < \zeta_c := (d' + d'/k)^{-1}$,

$d' = \max(d, 2)$, if $d \geq 1$ and $M_k(f) < \infty$ in [55] and with $\zeta = \zeta_c$ if furthermore $d \geq 3$ in [10].

To our knowledge, (1.9) and (1.10) are the first rates of convergence in MKW distance for the empirical law of large numbers associated to triangular array \mathcal{X}^N which coordinates are not i.i.d. random variables but only Kac's chaotic exchangeable random variables. The question of the optimality of the rates in (1.9) and (1.10) is an open (and we believe interesting) problem.

Now, following [17], we introduce the notion of *entropy chaos* and *Fisher information chaos* in the context of the ‘‘Kac's spheres’’ as follows. For any $j \in \mathbb{N}$, and $f, g \in \mathbf{P}(E^j)$, we define the usual relative entropy and usual relative Fisher information

$$H(f|g) := \frac{1}{j} \int_{E^j} u \log u g(dv), \quad I(f|g) := \frac{1}{j} \int_{E^j} \frac{|\nabla u|^2}{u} g(dv), \quad u := \frac{df}{dg},$$

where $u = \frac{df}{dg}$ stands for the Radon-Nikodym derivative of f with respect to g .

For $f \in \mathbf{P}(E)$ and $G^N \in \mathbf{P}_{sym}(\mathcal{KS}_N)$ such that $G_1^N \rightharpoonup f$ weakly in $\mathbf{P}(E)$, we say that (G^N) is

(a') f -entropy chaotic if $H(G^N|\sigma^N) \rightarrow H(f|\gamma)$, $H(f|\gamma) < \infty$;

(b') f -Fisher information chaotic if $I(G^N|\sigma^N) \rightarrow I(f|\gamma)$, $I(f|\gamma) < \infty$.

In a next step, we prove that for a large class of probability measures $f \in \mathbf{P}(\mathbb{R})$ the sequence (F^N) of associated ‘‘conditioned (to the Kac's spheres) product measures’’ is f -entropy chaotic as well as f -Fisher information chaotic, and we exhibit again rates for these convergences. The proof is mainly a careful rewriting and simplification of the proofs of the similar results (given without rate) in Theorems 9, 10, 19, 20 & 21 in [17].

We next generalize Theorem 1.4 to the Kac's spheres context. Additionally to the yet mentioned arguments, we use a general version of the HWI inequality proved by Lott and Villani in [46], see also [73, Theorem 30.21], and some entropy and Fisher inequalities on the Kac's spheres established by Carlen et al. [18] and improved by Barthe et al. [3].

All these results are motivated by the question of giving quantified strong version of propagation of chaos for Boltzmann-Kac jump model studied in [55] by Mouhot and the second author, where only quantitative uniform in time Kac's chaos is established. As a matter of fact, K. Carrapatoso in [19] extends the present analysis to the probability measures with support to the *Boltzmann's spheres* and proves a quantitative propagation result of entropy chaos.

Another outcome of our results is that we are able to give the following possible answer to [17, Open problem 11]:

Theorem 1.6. *Consider (G^N) a sequence of $\mathbf{P}_{sym}(\mathbb{R}^N)$ with support on the Kac's spheres \mathcal{KS}_N such that*

$$(1.12) \quad M_k(G_1^N) \leq C, \quad I(G^N|\sigma^N) \leq C,$$

for some $k \geq 2$ and $C > 0$. Also consider $f \in \mathbf{P}(\mathbb{R})$, satisfying $\int v^2 f(v) dv = 1$ and

$$(1.13) \quad f \geq \exp(-\alpha |v|^{k'} + \beta) \quad \text{on } \mathbb{R},$$

with $0 < k' < k$, $\alpha > 0$, $\beta \in \mathbb{R}$. If (G^N) is f -Kac's chaotic, then for any fixed $j \geq 1$, there holds

$$H(G_j^N|f^{\otimes j}) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

where $H(\cdot|\cdot)$ stands for the usual relative entropy functional defined in the flat space E^j . Remark that the boundedness of the k -th moment of G^N is useless when $k \leq 2$ (because

the support condition implies $M_2(G_1^N) = 1$) while the condition on the second moment of f is useless if $k > 2$ (because it is inherited from the properties of (G_1^N)).

Contrarily to the conditioned tensor product assumption made in [17, Theorem 9] which can be assumed at initial time for the stochastic Kac-Boltzmann process but which is not propagated along time, our assumptions (1.12) and (1.13) in Theorem 1.6, which may seem to be stronger in some sense, are in fact more natural since they are propagated along time. We refer to [55, 19] where such problems are studied.

Summary of Section 5. Here we investigate how our techniques can be used in the context of probability measures mixtures as introduced by De Finetti, Hewitt and Savage [24, 40] and general sequences of probability densities G^N of N undistinguishable particles as $N \rightarrow \infty$, without assuming chaos, as it is the case in [53, 13, 14] for instance. The results developed in that section are also used in a fundamental way in the recent work [32].

In a first step, we give a new proof of De Finetti, Hewitt and Savage theorem which is based on the use of the law of the empirical measure associated to the j first coordinates like in Diaconis and Freedman's proof [25] or Lions' proof [44], but where the compactness arguments are replaced by an argument of completeness. As a back product, we give a quantified equivalence of several notions of convergences of sequences of $\mathbf{P}_{sym}(E^N)$ to its possible mixture limit.

In a second step, we revisit the level 3 entropy and level 3 Fisher information theory for a probability measures mixture as developed since the work by Robinson and Ruelle [64] at least. We give a comprehensive and elementary proof of the fundamental result

$$(1.14) \quad \mathcal{K}(\pi) := \int_{\mathbf{P}(E)} K(\rho) \pi(d\rho) = \lim_{j \rightarrow \infty} \frac{1}{j} \int_{E^j} K(\pi_j)$$

for any probability measures mixture $\pi \in \mathbf{P}_k(\mathbf{P}(E))$, $k > 0$ (see paragraph 5.1 where the space $\mathbf{P}_k(\mathbf{P}(E))$ is defined), where π_j stands for the De Finetti, Hewitt and Savage projection of π on the j first coordinates and K stands for the Boltzmann's entropy or the Fisher information functional. It is worth noticing that while the representation formula (1.14) is well known when K stands for the Boltzmann's entropy, we believe that it is new when K stands for the Fisher information. The representation formula for the Fisher information is interesting for its own sake and it has also found an application as a key argument in the proof of propagation of chaos for system of vortices established in [32].

In our last result we establish a rate of convergence for the above limit (1.14) when K is the entropy functional mainly under a boundedness of the Fisher information hypothesis and we generalize such a quantitative result establishing links between several weak notions of convergence as well as strong (entropy) notion of convergence for sequences of probability densities $G^N \in \mathbf{P}_{sym}(E^N)$ as $N \rightarrow \infty$, without assuming chaos.

Acknowledgement. The authors would like to thank F. Bolley and C. Mouhot for many stimulating discussions about mean field limit and chaos, as well as N. Fournier for his suggestions that make possible to improve the statement and simplify the proof of the result on the level-3 Fisher information in section 5. The second author also would like to acknowledge I. Gentil and C. Villani for discussions about the HWI inequality and P.-L. Lions for discussions about entropy and mollifying tricks in infinite dimension. The second author acknowledges support from the project ANR-MADCOF.

2. KAC'S CHAOS

In this section we show the equivalence between several ways to measure Kac's chaos as stated in Theorem 1.2. We start presenting the framework we will deal with in the sequel, and thus making precise the definitions and notations used in the introductory section.

2.1. Definitions and notations. In all the sequel, we denote by E a closed subset of \mathbb{R}^d , $d \geq 1$, endowed with the usual topology, so that it is a locally compact Polish space. We denote by $\mathbf{P}(E)$ the space of probability measures on the Borel σ -algebra \mathcal{B}_E of E .

Monge-Kantorovich-Wasserstein (MKW) distances.

As they will be a cornerstone in that article, used in different setting, we briefly recall their definition and main properties, and refer to [72] for a very nice presentation.

On a general Polish space Z , for any distance $D : Z \times Z \rightarrow \mathbb{R}^+$ and $p \in [1, \infty)$, we define $W_{D,p}$ on $\mathbf{P}(Z) \times \mathbf{P}(Z)$ by setting for any $\rho_1, \rho_2 \in \mathbf{P}(Z)$

$$[W_{D,p}(\rho_1, \rho_2)]^p := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{Z \times Z} D(x, y)^p \pi(dx, dy)$$

where $\Pi(\rho_1, \rho_2)$ is the set of probability measures $\pi \in \mathbf{P}(Z \times Z)$ with first marginal ρ_1 and second marginal ρ_2 , that is $\pi(A \times Z) = \rho_1(A)$ and $\pi(Z \times A) = \rho_2(A)$ for any Borel set $A \subset Z$. It defines a distance on $\mathbf{P}(Z)$.

The phase spaces E^N (its marginal's space E^j) and $\mathbf{P}(E)$.

When we study system of N particles, the natural phase space is E^N . The space of marginals E^j for $1 \leq j \leq N$ are also important. We present here the different distances we shall use on these spaces.

- On E we will use mainly two distances :
 - the usual Euclidean distance denoted by $|x - y|$;
 - a bounded version of the square distance : $d_E(x, y) = |x - y| \wedge 1$ for any $x, y \in E$.
- On the space E^j for $1 \leq j$, we will also use the two distances
 - the *normalized* square distance $|X - Y|_2$ defined for any $X = (x_1, \dots, x_j) \in E^j$ and $Y = (y_1, \dots, y_j) \in E^j$ by

$$|X - Y|_2^2 := \frac{1}{j} \sum_{i=1}^j |x_i - y_i|^2;$$

- the *normalized* bounded distance $d_j = d_{E^j}$ defined by

$$(2.1) \quad d_{E^j}(X, Y) := \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i).$$

It is worth emphasizing that the normalizing factor $1/j$ is important in the sequel in order to obtain formulas independant of the number j of variables.

- The introduction of the empirical measures allows to “identify” our phase space E^N to a subspace of $\mathbf{P}(E)$. To be more precise, we denote by $\mathcal{P}_N(E)$ the set of empirical measures

$$\mathcal{P}_N(E) := \{\mu_X^N, X = (x_1, \dots, x_N) \in E^N\} \subset \mathbf{P}(E),$$

where μ_X^N stands for the empirical measure defined by (1.3) and associated to the configuration $X = (x_1, \dots, x_N) \in E^N$. We denote by $p_N : E^N \rightarrow \mathcal{P}_N(E)$ the application that maps a configuration to its empirical measure : $p_N(X) := \mu_X^N$.

- On our phase space $\mathbf{P}(E)$, we will use three different distances

- The usual MKW distance of order two W_2 defined as above with the choice $D(x, y) = |x - y|^2$

$$W_2(\rho_1, \rho_2)^2 = W_{|\cdot|_2, 2}(\rho_1, \rho_2)^2 := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{E \times E} |x - y|^2 \pi(dx, dy).$$

- The MKW distance W_1 associated to d_E defined by

$$(2.2) \quad W_1(\rho_1, \rho_2) = W_{d_E, 1}(\rho_1, \rho_2) := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{E \times E} d_E(x, y) \pi(dx, dy).$$

From the Kantorovich-Rubinstein duality theorem (see for instance [72, Theorem 1.14]) we have the following alternative characterization

$$(2.3) \quad \forall \rho_1, \rho_2 \in \mathbf{P}(E) \quad W_1(\rho_1, \rho_2) = \sup_{\|\varphi\|_{Lip} \leq 1} \int_E \varphi(x) (\rho_1(dx) - \rho_2(dx)),$$

where $\|\varphi\|_{Lip} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_E(x, y)}$ is the Lipschitz semi-norm relatively to the distance d_E . This semi-norm is closely related to the usual Lipschitz semi-norm since it satisfies

$$(2.4) \quad \frac{1}{2} (\|\nabla \varphi\|_\infty + \|\varphi - \varphi(0)\|_\infty) \leq \|\varphi\|_{Lip} \leq 2 (\|\nabla \varphi\|_\infty + \|\varphi\|_\infty) =: 2 \|\varphi\|_{W^{1, \infty}}.$$

It implies that W_1 is equivalent to the $(W^{1, \infty})'$ -distance, denoted by $D_{W^{1, \infty}}$,

$$D_{W^{1, \infty}}(\rho_1, \rho_2) := \sup_{\|\varphi\|_{W^{1, \infty}} \leq 1} \int_E \varphi(x) (\rho_1(dx) - \rho_2(dx)),$$

and more precisely

$$(2.5) \quad \frac{1}{2} D_{W^{1, \infty}} \leq W_1 \leq 2 D_{W^{1, \infty}}.$$

- The distance induced by the H^{-s} norm for $s > \frac{d}{2}$: for any $\rho, \eta \in \mathbf{P}(E)$

$$\|\rho - \eta\|_{H^{-s}}^2 := \int_{\mathbb{R}^d} |\hat{\rho}(\xi) - \hat{\eta}(\xi)|^2 \frac{d\xi}{\langle \xi \rangle^{2s}}$$

where $\hat{\rho}$ denotes the Fourier transform of ρ (which may always be seen as a measure on the whole \mathbb{R}^d), and $\langle \xi \rangle = \sqrt{1 + |\xi|^2}$.

- We will often restrict ourself to the spaces $\mathbf{P}_k(E)$ of probability measures with finite moment of order $k > 0$ defined by

$$\mathbf{P}_k(E) := \{\rho \in \mathbf{P}(E) \text{ s.t. } M_k(\rho) := \int_E \langle v \rangle^k \rho(dv) < +\infty\}.$$

The probability measures space $\mathbf{P}(E^N)$, its marginals spaces $\mathbf{P}(E^j)$, and $\mathbf{P}(\mathbf{P}(E))$.
The next step is to consider probability measures on the configuration spaces.

- The space $\mathbf{P}(E^N)$ will be endowed with two distances
 - W_1 the MKW distance on $\mathbf{P}(E^N)$ associated to d_{E^N} and $p = 1$, which has the same properties as the one constructed on $\mathbf{P}(E)$ and satisfies in particular the Kantorovich-Rubinstein formulation (2.3).
 - W_2 the MKW distance associated to the normalized square distance $|\cdot|_2$ defined above.

Remark that we will only work on the subspace $\mathbf{P}_{sym}(E^N)$ of borelian probability measures which are invariant under coordinates permutations.

• On the probability measures space $\mathbf{P}(\mathbf{P}(E))$, we can define different distances thanks to the Monge-Kantorovich-Wasserstein construction. We will use three of them

– \mathcal{W}_1 , the MKW distance induced by the cost function W_1 on $\mathbf{P}(E)$. In short

$$(2.6) \quad \mathcal{W}_1(\alpha_1, \alpha_2) = W_{W_1,1}(\alpha_1, \alpha_2) := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho_1, \rho_2) \pi(d\rho_1, d\rho_2),$$

– \mathcal{W}_2 , the MKW distance induced by the cost function W_2^2 on $\mathbf{P}(E)$. In short

$$\mathcal{W}_2(\alpha_1, \alpha_2)^2 = W_{W_2,2}(\alpha_1, \alpha_2)^2 := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_2^2(\rho_1, \rho_2) \pi(d\rho_1, d\rho_2),$$

– $\mathcal{W}_{H^{-s}}$, the MKW distance induced by the cost function $\|\cdot\|_{H^{-s}}^2$ on $\mathbf{P}(E)$. In short

$$\mathcal{W}_{H^{-s}}(\alpha_1, \alpha_2)^2 = \mathcal{W}_{\|\cdot\|_{H^{-s}},2}(\alpha_1, \alpha_2)^2 := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho_1 - \rho_2\|_{H^{-s}}^2 \pi(d\rho_1, d\rho_2).$$

• Remark that the application "empirical measure" p_N allows to define by push-forward a canonical map between $\mathbf{P}(E^N)$ and $\mathbf{P}(\mathbf{P}(E))$. For $G^N \in \mathbf{P}(E^N)$ we denote its image under the application p_N by $\hat{G}^N \in \mathbf{P}(\mathbf{P}(E)) : \hat{G}^N := G^N_{\#} p_N$. In other words, \hat{G}^N is the unique probability measure in $\mathbf{P}(\mathbf{P}(E))$ which satisfies the duality relation

$$(2.7) \quad \forall \Phi \in C_b(\mathbf{P}(E)) \quad \langle \hat{G}^N, \Phi \rangle = \int_{E^N} \Phi(\mu_X^N) G^N(dX).$$

More properties of the space $\mathbf{P}(\mathbf{P}(E))$.

• **Marginals of probability measures on $\mathbf{P}(\mathbf{P}(E))$.** We can define a mapping from $\mathbf{P}(\mathbf{P}(E))$ onto $\mathbf{P}(E^j)$ in the following way. For any $\alpha \in \mathbf{P}(\mathbf{P}(E))$ we define the projection $\alpha_j \in \mathbf{P}(E^j)$ thanks to the relation

$$\alpha_j := \int \rho^{\otimes j} d\alpha(\rho).$$

It may also be restated using *polynomial fonctions* : for any $\varphi \in C_b(E^j)$ we define the monomial (of order j) function $R_\varphi \in C_b(\mathbf{P}(E))$ by

$$(2.8) \quad \forall \rho \in \mathbf{P}(E) \quad R_\varphi(\rho) := \int_{E^j} \varphi(X) \rho^{\otimes j}(dX).$$

We remark that the monomial functions of all orders generate an algebra of continuous function (for the weak convergence of measures) that are called polynomials. When E is compact so that $\mathbf{P}(E)$ is also compact, they form a dense subset of $C_b(\mathbf{P}(E))$ thanks to the Stone-Weierstrass theorem.

In terms of polynomial functions, the marginal α_j may be defined by

$$\forall \varphi \in C_b(E^j) \quad \langle \alpha_j, \varphi \rangle := \langle \alpha, R_\varphi \rangle.$$

• Starting from $G^N \in \mathbf{P}_{sym}(E^N)$, we can define its push-forward \hat{G}^N and then for any $1 \leq j \leq N$ the marginals of the push-forward $\hat{G}_j^N := (\hat{G}^N)_j \in \mathbf{P}_{sym}(E^j)$. They satisfy the duality relation

$$(2.9) \quad \forall \varphi \in C_b(E^j) \quad \langle \hat{G}_j^N, \varphi \rangle := \int_{E^N} R_\varphi(\mu_X^N) G^N(dX).$$

We emphasize that it is not equal to G_j^N the j -th marginal of G^N , but we will see later that the two probability measures G_j^N and \hat{G}_j^N are close (a precise version is recalled in Lemma 2.8).

Different quantities measuring chaoticity. Now that everything has been defined, we introduce the quantities that we will use to quantify the chaoticity of a sequence $G^N \in \mathbf{P}_{sym}(E^N)$ of symmetric probability measure with respect to a profil $f \in \mathbf{P}(E)$:

- The chaoticity can be mesured on E^j for $j \geq 2$. For any $1 \leq j \leq N$, we set

$$\Omega_j(G^N; f) := W_1(G_j^N, f^{\otimes j}),$$

- and also on $\mathbf{P}(E)$ by

$$\Omega_\infty(G^N; f) := \mathcal{W}_1(\hat{G}^N, \delta_f) = \int_{E^N} W_1(\mu_X^N, f) G^N(dX),$$

since there is only one transference plan $\alpha \otimes \delta_f$ in $\Pi(\alpha, \delta_f)$.

2.2. Equivalence of distances on $\mathbf{P}(E)$, $\mathbf{P}_{sym}(E^N)$ and $\mathbf{P}(\mathbf{P}(E))$.

To quantify the equivalence between the distances defined above on $\mathbf{P}(E)$, we will need some assumption on the moments. The metrics W_1 , W_2 and $\|\cdot\|_{H^{-s}}$ are uniformly topologically equivalent in $\mathbf{P}_k(E)$ for any $k > 0$. More precisely, we have

Lemma 2.1. *Choose $f, g \in \mathbf{P}(E)$. For any $k > 0$, denote $\mathcal{M}_k := M_k(f) + M_k(g)$.*

(i) *For any $k > 0$ and $s \geq 1$, there exists $C := C(d) \left[1 + \left(\frac{s-1}{2}\right)^{\frac{s-1}{2}}\right]$, such that there holds*

$$(2.10) \quad W_1(f, g) \leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \|f - g\|_{H^{-s}}^{\frac{2k}{d+2ks}}.$$

(ii) *For any $k > 2$, there holds*

$$(2.11) \quad W_2(f, g) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{1/k} W_1(f, g)^{1/2-1/k}.$$

(iii) *Without moment assumptions and for any $s > \frac{d+1}{2}$, there exists a constant $C = C(s, d)$ such that there holds*

$$W_1(f, g) \leq W_2(f, g), \quad \|f - g\|_{H^{-s}} \leq C W_1(f, g)^{\frac{1}{2}}.$$

We remark that we have kept the explicit dependance on s of the constant appearing in (i) in order to be able to perform some optimization on s later. The important point is that the constant may be choosen independant of s if s varies in a compact set.

PROOF OF LEMMA 2.1. The proof is a mere adaptation of classical results on comparison of distances in probability measures spaces as it can be found in [62, 21, 55] for instance. We nevertheless sketch it for the sake of completeness.

Proof of i). We consider a truncation sequence $\chi_R(x) = \chi(x/R)$, $R > 0$, with $\chi \in C_c^\infty(\mathbb{R}^d)$, $\|\nabla \chi\|_\infty \leq 1$, $0 \leq \chi \leq 1$, $\chi \equiv 1$ on $B(0, 1)$, and the sequence of mollifiers $\gamma_\varepsilon(x) = \varepsilon^{-d} \gamma(x/\varepsilon)$, $\varepsilon > 0$, with $\gamma(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$, so that $\hat{\gamma}_\varepsilon(\xi) = \exp(-\varepsilon^2 |\xi|^2/2)$. In view of the equivalence of distance (2.5), we choose a $\varphi \in W^{1,\infty}(\mathbb{R}^d)$ such that $\|\varphi\|_{W^{1,\infty}} \leq 1$, we define $\varphi_R := \varphi \chi_R$, $\varphi_{R,\varepsilon} = \varphi_R * \gamma_\varepsilon$ and we write

$$\int \varphi (df - dg) = \int \varphi_{R,\varepsilon} (df - dg) + \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) + \int (\varphi - \varphi_R) (df - dg).$$

For the last term, we have

$$\forall R > 0 \quad \left| \int (\varphi_R - \varphi) (df - dg) \right| \leq \int_{B_R^c} \|\varphi\|_\infty \frac{|x|^k}{R^k} (df + dg) \leq \frac{M_k}{R^k}.$$

For the second term, we observe that

$$\|\varphi_R - \varphi_{R,\varepsilon}\|_\infty \leq \|\nabla \varphi_R\|_\infty \int_{\mathbb{R}^d} \gamma_\varepsilon(x) |x| dx \leq C(d) \varepsilon,$$

and we get

$$\left| \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) \right| \leq C(d) \varepsilon.$$

Finally, the first term can be estimated by

$$\left| \int \varphi_{R,\varepsilon} (df - dg) \right| \leq \|\varphi_{R,\varepsilon}\|_{H^s} \|f - g\|_{H^{-s}},$$

with for any $R \geq 1$ and $\varepsilon \in (0, 1]$

$$\begin{aligned} \|\varphi_{R,\varepsilon}\|_{H^s} &= \left(\int \langle \xi \rangle^2 |\widehat{\varphi \chi_R}|^2 \langle \xi \rangle^{2(s-1)} |\hat{\gamma}_\varepsilon|^2 d\xi \right)^{1/2} \\ &\leq \|\varphi \chi_R\|_{H^1} \|\langle \xi \rangle^{s-1} \hat{\gamma}_\varepsilon(\xi)\|_{L^\infty} \leq C(d) R^{d/2} \|\langle \xi \rangle^{s-1} \hat{\gamma}_\varepsilon(\xi)\|_{L^\infty} \end{aligned}$$

The infinite norm is finite and a simple optimization leads to

$$\|\langle \xi \rangle^{s-1} \hat{\gamma}_\varepsilon(\xi)\|_{L^\infty} \leq \left(\frac{s-1}{2} \right)^{\frac{s-1}{2}} \varepsilon^{-(s-1)_+},$$

with the natural convention $0^0 = 1$. All in all, we have

$$W_1(f, g) \leq C(d) \left[1 + \left(\frac{s-1}{2} \right)^{\frac{s-1}{2}} \right] \left(\varepsilon + \frac{M_k}{R^k} + R^{\frac{d}{2}} \varepsilon^{-(s-1)} \|f - g\|_{H^{-s}} \right).$$

This yields to (2.10) by optimizing the paramater ε and R with

$$R = M_k^{\frac{2s}{d+2ks}} \|f - g\|_{H^{-s}}^{-\frac{2}{d+2ks}}, \quad \text{and} \quad \varepsilon = M_k^{\frac{d}{d+2ks}} \|f - g\|_{H^{-s}}^{\frac{2k}{d+2ks}}.$$

Proof of ii). We have for any $R \geq 1$ the inequality

$$\forall x, y \in E, \quad |x - y|^2 \leq R^2 d_E(x, y) + \frac{2^k}{R^{k-2}} (|x|^k + |y|^k)$$

from which we deduce

$$\begin{aligned} W_2(f, g)^2 &\leq R^2 \inf_{\pi \in \Pi(f, g)} \int_{E \times E} d_E(x, y) \pi(dx, dy) \\ &\quad + \frac{2^k}{R^{k-2}} \sup_{\pi \in \Pi(f, g)} \int_{E \times E} (|x_i|^k + |y_i|^k) \pi(dx, dy) \\ &\leq R^2 W_1(f, g) + \frac{2^k}{R^{k-2}} (M_k(f) + M_k(g)), \end{aligned}$$

and then we get with $(R/2)^k = \mathcal{M}_k/W_1$

$$(2.12) \quad W_2(f, g) \leq 2^{3/2} \mathcal{M}_k^{1/k} W_1(f, g)^{1/2-1/k}.$$

Proof of iii). The first point is classical. The second relies on the fact that $\|\delta_x - \delta_y\|_{H^{-s}}^2 \leq C d_E(x, y)$. □

There is also a similar result on E^N , where the H^{-s} norm is less usefull.

Lemma 2.2. Choose $F^N, G^N \in \mathbf{P}_{\text{sym}}(E^N)$. For any $k > 0$, denotes

$$\mathcal{M}_k := M_k(F_1^N) + M_k(G_1^N).$$

For any $k > 2$, it holds that

$$(2.13) \quad W_2(F^N, G^N) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{1/k} W_1(F^N, G^N)^{1/2-1/k}.$$

It also holds without moment assumptions that $W_1(F^N, G^N) \leq W_2(F^N, G^N)$.

PROOF OF LEMMA 2.2. The proof is a simple generalization of (2.11) to the case of N variables. We skip it. \square

The inequalities of Lemma 2.1 also sum well on $\mathbf{P}(\mathbf{P}(E))$ in order to get

Lemma 2.3. Choose $\alpha, \beta \in \mathbf{P}(\mathbf{P}(E))$, and for $k > 0$ define

$$\mathcal{M}_k := M_k(\alpha) + M_k(\beta) := \int M_k(\rho) [\alpha + \beta](d\rho) = M_k(\alpha_1) + M_k(\beta_1).$$

(i) For any $s \geq 1$ and with the same constant $C(d, s)$ as in point (i) of Lemma 2.1 we have for any $k > 0$,

$$(2.14) \quad \mathcal{W}_1(\alpha, \beta) \leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\alpha, \beta)^{\frac{2k}{d+2ks}}.$$

ii) For any $k > 2$, it also holds

$$(2.15) \quad (ii) \quad \mathcal{W}_2(\alpha, \beta) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{\frac{1}{k}} \mathcal{W}_1(\alpha, \beta)^{\frac{1}{2}-\frac{1}{k}},$$

iii) It holds without moment assumption that $\mathcal{W}_1 \leq \mathcal{W}_2$ and $\mathcal{W}_{H^{-s}} \leq C \mathcal{W}_1^{\frac{1}{2}}$ for $s > \frac{d+1}{2}$ with a constant $C = C(s, d)$.

PROOF OF LEMMA 2.3. All the above estimates are simple summations of the corresponding estimate of Lemma 2.1. We only prove i).

$$\begin{aligned} \mathcal{W}_1(\alpha, \beta) &= \inf_{\Pi \in \Pi(\alpha, \beta)} \int W_1(\rho, \eta) \Pi(d\rho, d\eta) \\ &\leq C \inf_{\Pi \in \Pi(\alpha, \beta)} \int [M_k(\rho) + M_k(\eta)]^{\frac{d}{d+2ks}} \|\rho - \eta\|_{H^{-s}}^{\frac{2k}{d+2ks}} \Pi(d\rho, d\eta) \\ &\leq C \left(\int M_k(\rho) [\alpha + \beta](d\rho) \right)^{\frac{d}{d+2ks}} \left(\inf_{\Pi \in \Pi(\alpha, \beta)} \int \|\rho - \eta\|_{H^{-s}}^{\frac{1}{s}} \Pi(d\rho, d\eta) \right)^{\frac{2ks}{d+2ks}} \\ &\leq C [M_k(\alpha) + M_k(\beta)]^{\frac{d}{d+2ks}} \left(\inf_{\Pi \in \Pi(\alpha, \beta)} \int \|\rho - \eta\|_{H^{-s}}^2 \Pi(d\rho, d\eta) \right)^{\frac{k}{d+2ks}} \\ &\leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\alpha, \beta)^{\frac{2k}{d+2ks}} \end{aligned}$$

where we have successively used the inequality (2.10), Hölder inequality, the definition of the moment of α and β , and Jensen inequality. \square

2.3. Quantified equivalence of chaos. This section is devoted to the proof of Theorem 1.2, or more precisely, to the proof of the following accurate version of Theorem 1.2.

Theorem 2.4. *For any $G^N \in \mathbf{P}_{\text{sym}}(E^N)$ and $f \in \mathbf{P}(E)$, there holds*

$$(2.16) \quad (i) \quad \forall 1 \leq j \leq \ell \leq N \quad \Omega_j(G^N; f) \leq 2\Omega_\ell(G^N; f),$$

$$(2.17) \quad (ii) \quad \forall 1 \leq j \leq N \quad \Omega_j(G^N; f) \leq \Omega_\infty(G^N; f) + \frac{j^2}{N}.$$

For any $k > 0$ and any $0 < \gamma < \frac{1}{d+1+\frac{d}{k}}$, there exists a explicit constant $C := C(d, \gamma, k)$ such that

$$(2.18) \quad (iii) \quad \Omega_\infty(G^N; f) \leq C \mathcal{M}_k^{\frac{1}{k}} \left(\Omega_2(G^N; f) + \frac{1}{N} \right)^\gamma,$$

where as usual $\mathcal{M}_k := M_k(f) + M_k(G_1^N)$.

For any $k > 0$ and any $0 < \gamma < \frac{1}{d'+\frac{d'}{k}}$, with $d' = \max(d, 2)$, there exists a constant $C := C(d, \gamma, k)$ such that

$$(2.19) \quad (iv) \quad |\Omega_N(G^N; f) - \Omega_\infty(G^N; f)| \leq C \frac{M_k(f)^{1/k}}{N^\gamma}.$$

Let us make some remarks about the above statement. Roughly speaking, the two first inequalities are in the good sense: the measure of chaos for a certain number of particles is bounded by the measure of chaos with more particles, and even in the sense of empirical measure (i.e. with Ω_∞). Let us however observe that the second inequality is meaningful only when the number j of particles in the left hand side is not too high, typically $j = o(\sqrt{N})$. The third inequality is in the "bad sense" and it is maybe the most important one, since it provides an estimate of the measure of chaos in the sense of empirical measures by the measure of chaos for two particles only. It is for instance a key ingredient in [55]. See also corollary 2.11 for versions adapted to probability measures with compact support or with exponential moment. The last inequality compares the measure of chaos at N particles to its measure in the sense of empirical distribution. It seems new and it will be a key argument in the next sections in order to make links between the Kac's chaos, the entropy chaos and the Fisher information chaos.

Remark 2.5. *In the inequality (2.18), the Ω_2 term in the right hand side may be replaced by any Ω_ℓ for $\ell \geq 2$, but it cannot be replaced by Ω_1 , which does not measures chaoticity, as it is well known. We give a counter-example for the sake of completeness. We choose g and h two distinct probability measures on E , and take $f := \frac{1}{2}(g + h)$. We consider the probability measure $G \in \mathbf{P}(\mathbf{P}(E))$, and its associated sequence (G^N) of marginal probability measures on $\mathbf{P}(E^N)$ defined by*

$$G = \frac{1}{2}(\delta_g + \delta_h), \quad G^N := \frac{1}{2}g^{\otimes N} + \frac{1}{2}h^{\otimes N}.$$

As $G_1 = f$, $\Omega_1(G^N, f) = 0$ for all N , inequality (2.18) with Ω_2 replaced by Ω_1 will imply that $\Omega_\infty(G^N, f)$ goes to zero. But from inequality (2.17) of Theorem 2.4

$$W_1(G^2, f^{\otimes 2}) = \Omega_2(G^N, f) \leq \Omega_\infty(G^N, f) + \frac{C}{N}.$$

There is a contradiction since $G^2 \neq f^{\otimes 2}$ except if $g = h$.

We begin with some probably well known elementary inequalities and identities concerning Monge-Kantorovich-Wasserstein distances in space product. For the sake of completeness we will nevertheless sketch the proofs of them. Remark that the two first formulas are particularly simple thanks to the choice of the normalization (2.1), and that they remains valid if we replace d_j by the normalized l^1 -distance $\frac{1}{j} \sum_i |x_i - y_i|$.

Proposition 2.6. *a) - For any $F^N, G^N \in \mathbf{P}_{\text{sym}}(E^N)$ and $1 \leq j \leq N$, there holds*

$$(2.20) \quad W_1(F_j^N, G_j^N) \leq \left(\frac{j}{N} \left\lceil \frac{N}{j} \right\rceil \right)^{-1} W_1(F^N, G^N) \leq 2 W_1(F^N, G^N).$$

b) - For any $f, g \in P(E)$, there holds

$$(2.21) \quad W_1(f^{\otimes N}, g^{\otimes N}) = W_1(f, g).$$

c) - For any $f, g, h \in P(E)$, there holds

$$(2.22) \quad 2 W_1(f \otimes h, g \otimes h) = W_1(f, g).$$

As a immediate corollary of (2.20) with $N := \ell$, $F^\ell := f^{\otimes \ell}$ and $G^\ell := G_\ell^N$, we obtain the first inequality (2.16) of Theorem 2.4.

As can be seen in the following proof, similar results also holds for MKW distances constructed with arbitrary distance D and exponents p , and therefore for the W_2 distance. We do not state them precisely, but they will be useful in the proof of the next Lemma 2.7.

PROOF OF PROPOSITION 2.6.

Proof of (2.20). Consider $\pi \in \Pi(F^N, G^N)$ an optimal transference plan in (2.2). Introducing the Euclidean division, $N = nj + r$, $0 \leq r \leq j - 1$, and writing $X = (X_1, \dots, X_n, X_0) \in E^N$, $Y = (Y_1, \dots, Y_n, Y_0) \in E^N$, with $X_i, Y_i \in E^j$, $1 \leq i \leq n$, $X_0, Y_0 \in E^r$, we have

$$\begin{aligned} W_1(F^N, G^N) &= \int_{E^{2N}} d_{E^N}(X, Y) \pi(dX, dY) \\ &= \frac{1}{N} \int_{E^{2N}} \left(\sum_{i=1}^n j d_{E^j}(X_i, Y_i) + r d_{E^r}(X_0, Y_0) \right) \pi(dX, dY) \\ &\geq \frac{j}{N} \sum_{i=1}^n \int_{E^{2j}} d_{E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dY_i), \end{aligned}$$

with $\tilde{\pi}_i \in \Pi(\tilde{F}_i, \tilde{G}_i)$, where \tilde{F}_i and $\tilde{G}_i \in \mathbf{P}(E^j)$ denote the marginal probability measures of F^N and G^N on the i -th block of variables. From the symmetry hypothesis, we have $\tilde{F}_i = \tilde{F}_1 = F_j^N$ and $\tilde{G}_i = \tilde{G}_1 = G_j^N$ for any $1 \leq i \leq n$. As a consequence, we have

$$\int_{E^{2j}} d_{E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dY_i) \geq W_1(F_j^N, G_j^N),$$

and we then deduce the first inequality in (2.20). Since the integer portion $n := [N/j]$ is larger than 1, we have

$$\frac{j}{N} \left\lceil \frac{N}{j} \right\rceil = \frac{nj}{nj+r} \geq \frac{nj}{nj+j} \geq \frac{1}{2},$$

from which we deduce the second inequality in (2.20).

Proof of (2.21). We consider $\alpha \in \Pi(f, g)$ an optimal transference plan for the $W_1(f, g)$ distance and we define the associated transference plan $\bar{\pi} := \alpha^{\otimes N} \in \Pi(f^{\otimes N}, g^{\otimes N})$ by

$$\forall A_i, B_i \in E \quad \bar{\pi}(A_1 \times \dots \times A_N \times B_1 \times \dots \times B_N) = \alpha(A_1 \times B_1) \times \dots \times \alpha(A_N \times B_N).$$

By definition of $W_1(f^{\otimes N}, g^{\otimes N})$, we then have

$$W_1(f^{\otimes N}, g^{\otimes N}) \leq \frac{1}{N} \sum_{i=1}^N \int_{E^{2N}} d(x_i, y_i) \bar{\pi}(dX, dY) = W_1(f, g).$$

Since the first inequality in (2.20) in the case $j = 1$ implies the reverse inequality, the above inequality is an equality.

Proof of (2.22). On the one hand, from the definition of the distance W_1 by transference plans, we have for an optimal transference plan $\pi \in \Pi(f \otimes h, g \otimes h)$ the inequality

$$\begin{aligned} W_1(f \otimes h, g \otimes h) &= \frac{1}{2} \int_{E^4} (d_E(x_1, y_1) + d_E(x_2, y_2)) \pi(dx_1, dx_2, dy_1, dy_2) \\ &\geq \frac{1}{2} \int_{E^4} d_E(x_1, y_1) \pi_1(dx_1, dy_1) \geq \frac{1}{2} W_1(f, g), \end{aligned}$$

since the 1-marginal π_1 defined by $\pi_1(A \times B) = \pi(A \times E \times B \times E)$ for any $A, B \in \mathcal{B}_E$ belongs to the transference plans set $\Pi(f, g)$.

On the other hand, considering an optimal transference plan $\pi \in \Pi(f, g)$ for the W_1 distance, we define the associated transference plan $\bar{\pi}(dx, dy) := \pi(dx_1, dy_1) \otimes h(dx_2) \delta_{y_2=x_2} \in \Pi(f \otimes h, g \otimes h)$, and we observe that

$$\begin{aligned} W_1(f \otimes h, g \otimes h) &\leq \frac{1}{2} \int_{E^4} (d_E(x_1, y_1) + d_E(x_2, y_2)) \bar{\pi}(dx_1, dx_2, dy_1, dy_2) \\ &= \frac{1}{2} \int_{E^4} d_E(x_1, y_1) \pi(dx_1, dy_1) = \frac{1}{2} W_1(f, g). \end{aligned}$$

We obtain (2.22) by gathering these two inequalities. \square

We next prove another lemma that allows to compare a distance between measures on $\mathbf{P}(\mathbf{P}(E))$ and a distance between their marginals on E^j , and thus to compare Ω_ℓ and Ω_∞ .

Lemma 2.7. *For any distance D on E and $p \geq 1$, extend D on E^j with $D_{j,p}(V, W)^p = \frac{1}{j} \sum_i D(v_i, w_i)^p$, and define the associated MKW distance $W_{D_{j,p},p}$ on $\mathbf{P}(E^j)$ and the MKW distance $\mathcal{W}_{W_D,p}$ on $\mathbf{P}(\mathbf{P}(E))$ associated to W_D and p . Let α and β be two probability measures on $\mathbf{P}(\mathbf{P}(E))$. Then, for any $j \in \mathbb{N}$,*

$$(2.23) \quad W_{D_{j,p},p}(\alpha_j, \beta_j) \leq \mathcal{W}_{W_D,p}(\alpha, \beta)$$

That is in particular true for the MKW distances W_1 and W_2 defined in section 2.1

$$\forall j \in \mathbb{N}, \quad W_2(\alpha_j, \beta_j) \leq \mathcal{W}_2(\alpha, \beta), \quad W_1(\alpha_j, \beta_j) \leq \mathcal{W}_1(\alpha, \beta).$$

PROOF OF LEMMA 2.7. For simplicity we denote for any j , $W_{D_{j,p},p} = W_D$. We choose any transference plan Π between α and β and write

$$\begin{aligned} [W_D(\alpha_j, \beta_j)]^p &= \left[W_D \left(\int \rho^{\otimes j} \alpha(d\rho), \int \rho^{\otimes j} \beta(d\rho) \right) \right]^p \\ &= \left[W_D \left(\int \rho^{\otimes j} \pi(d\rho, d\eta), \int \eta^{\otimes j} \pi(d\rho, d\eta) \right) \right]^p \\ &\leq \left[\int W_D(\rho^{\otimes j}, \eta^{\otimes j}) \pi(d\rho, d\eta) \right]^p \\ &\leq \int [W_D(\rho, \eta)]^p \pi(d\rho, d\eta), \end{aligned}$$

where we have used the convexity property of the Wasserstein distance, the equivalent of equality (2.21) in our general case, and Jensen inequality. By optimisation on π we obtain the claimed inequality. \square

As a consequence of a classical combinatory trick, which goes back at least to [36], we have

Lemma 2.8 (Quantification of the equivalence $G_j^N \sim \hat{G}_j^N$). *For any $G^N \in \mathbf{P}_{sym}(E^N)$ and any $1 \leq j \leq 1 + N/2$, we have*

$$\|G_j^N - \hat{G}_j^N\|_{TV} \leq 2 \frac{j(j-1)}{N} \quad \text{and} \quad W_1(G_j^N, \hat{G}_j^N) = \frac{j(j-1)}{N},$$

and in particular the first marginals are equal: $G_1^N = \hat{G}_1^N$.

PROOF OF LEMMA 2.8. The second inequality is a straightforward consequence of the first inequality together with the use of

$$W_1(G_j^N, \hat{G}_j^N) \leq \frac{1}{2} \|G_j^N - \hat{G}_j^N\|_{TV}.$$

A proof of the later may be found in [72, Proposition 7.10], in a slightly different context. Here, the better factor 1/2 can be obtained because of the stronger assumptions of our setting (the distance d_{E^j} we deal with here is bounded by 1).

The first inequality is a simple and classical combinatorial computation, see for instance [36], [67, Proposition 2.2], [57, Lemma 4.2] or [55, Lemma 3.3]. We briefly sketch the proof for the convenience of the reader.

For $1 \leq j \leq N$, we denote by \mathcal{C}_j^N the set of maps from $\{1, \dots, j\}$ into $\{1, \dots, N\}$, and by \mathcal{A}_j^N the subset of \mathcal{C}_j^N made of the one-to-one maps. Remark that we have

$$|\mathcal{C}_j^N| = N^j, \quad |\mathcal{A}_j^N| = \frac{N!}{(N-j)!}.$$

Thanks to the symmetry assumption made on G^N , we may write for any $\varphi \in C_b(E^j)$

$$\langle G_j^N, \varphi \rangle = \int_{E^N} \varphi(x_1, \dots, x_j) G^N(dX) = \frac{(N-j)!}{N!} \sum_{s \in \mathcal{A}_j^N} \int_{E^N} \varphi(x_{s(1)}, \dots, x_{s(j)}) G^N(dX)$$

From the definition of \hat{G}_j^N we also get

$$\begin{aligned} \langle \hat{G}_j^N, \varphi \rangle &= \int_{\mathbf{P}(E)} \left(\int \varphi(y_1, \dots, y_j) \rho^{\otimes j}(dY^j) \right) \hat{G}^N(d\rho) \\ &= \int_{E^N} \left(\int \varphi(y_1, \dots, y_j) (\mu_X^N)^{\otimes j}(dY^j) \right) G^N(dX^N) \\ &= \frac{1}{N^j} \sum_{s \in \mathcal{C}_j^N} \int_{E^N} \varphi(x_{s(1)}, \dots, x_{s(j)}) G^N(dX^N). \end{aligned}$$

The difference is then equals to

$$\begin{aligned} \langle G_j^N - \hat{G}_j^N, \varphi \rangle &= \left(\frac{(N-j)!}{N!} - \frac{1}{N^j} \right) \sum_{s \in \mathcal{A}_j^N} \int_{E^N} \varphi(x_{s(1)}, \dots, x_{s(j)}) G^N(dX) \\ &\quad - \frac{1}{N^j} \sum_{s \in \mathcal{C}_j^N \setminus \mathcal{A}_j^N} \int_{E^N} \varphi(x_{s(1)}, \dots, x_{s(j)}) G^N(dX^N) \end{aligned}$$

and may be bounded by

$$\begin{aligned} |\langle G_j^N - \hat{G}_j^N, \varphi \rangle| &\leq \left(1 - \frac{N!}{N^j(N-j)!} \right) \|\varphi\|_{L^\infty} + \frac{1}{N^j} |\mathcal{C}_j^N \setminus \mathcal{A}_j^N| \|\varphi\|_{L^\infty} \\ &= 2 \left(1 - \frac{N!}{N^j(N-j)!} \right) \|\varphi\|_{L^\infty}. \end{aligned}$$

For $N \geq 2(j-1)$, we can bound the right hand side thanks to

$$\begin{aligned} 1 - \frac{N!}{(N-j)! N^j} &= 1 - \left(1 - \frac{1}{N} \right) \cdots \left(1 - \frac{j-1}{N} \right) = 1 - \exp \left(\sum_{i=0}^{j-1} \ln \left(1 - \frac{i}{N} \right) \right) \\ &\leq 1 - \exp \left(-2 \sum_{i=0}^{j-1} \frac{i}{N} \right) \leq 2 \sum_{i=0}^{j-1} \frac{i}{N} \leq \frac{j(j-1)}{N}, \end{aligned}$$

where we have used

$$\forall x \in [0, 1/2], \quad \ln(1-x) \geq -2x \quad \text{and} \quad \forall x \in \mathbb{R}, \quad e^{-x} \geq 1-x.$$

We eventually get for $j \leq 1 + N/2$

$$\|G_j^N - \hat{G}_j^N\|_{TV} = \sup_{\|\varphi\|_\infty \leq 1} \langle G_j^N - \hat{G}_j^N, \varphi \rangle \leq 2 \frac{j(j-1)}{N},$$

which ends the proof. \square

Applying the previous lemmas 2.7 and 2.8, we can bound Ω_j by Ω_∞ and some rest. This is the second inequality (2.17) of theorem 2.4.

PROOF OF INEQUALITY (2.17) IN THEOREM 2.4. We simply write

$$\begin{aligned} \Omega_j(G^N, f) = W_1(G_j^N, f^{\otimes j}) &\leq W_1(G_j^N, \hat{G}_j^N) + W_1(\hat{G}_j^N, f^{\otimes j}) \\ &\leq \frac{j^2}{N} + \mathcal{W}_1(\hat{G}^N, \delta_f) = \frac{j^2}{N} + \Omega_\infty(G^N, f), \end{aligned}$$

thanks to the two previous lemmas 2.7 and 2.8. \square

We establish now the key estimate which will lead to the third inequality (2.18) in Theorem 2.4 where Ω_∞ is controlled by Ω_2 . Following [55, Lemma 4.2], the main idea is to use as an intermediate step the H^{-s} norm on $\mathbf{P}(E)$, rather than the Wassertsein W_1 distance, because it is a monomial function of order two on $\mathbf{P}(E)$, and thus has a nice algebraic structure. This fact is stated in the following elementary lemma.

Lemma 2.9. *For $s > d/2$, define $\Phi_s : \mathbb{R}^d \rightarrow \mathbb{R}$ by*

$$(2.24) \quad \forall z \in \mathbb{R}^d, \quad \Phi_s(z) := \int_{\mathbb{R}^d} e^{-iz \cdot \xi} \frac{d\xi}{\langle \xi \rangle^{2s}}.$$

The function Φ_s is radial, bounded, and furthermore if $s > \frac{d+1}{2}$, it is Lipschitz. For any $\rho, \eta \in \mathbf{P}(E)$

$$(2.25) \quad \|\rho - \eta\|_{H^{-s}}^2 = \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes \eta)(dx, dy) + \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\eta^{\otimes 2} - \eta \otimes \rho)(dx, dy),$$

and for any $\rho \in \mathbf{P}(E)$

$$\|\rho\|_{H^{-s}}^2 = \int_{\mathbb{R}^{2d}} \Phi(x-y) \rho^{\otimes 2}(dx, dy),$$

which means that the norm H^{-s} on $\mathbf{P}(E)$ is the monomial function of order two associated to the function $(x, y) \mapsto \Phi_s(x-y)$.

PROOF OF LEMMA 2.9. We obtain that Φ_s is bounded from the fact that $\int_{\mathbb{R}^d} \langle \xi \rangle^{-2s} d\xi$ is finite for $s > d/2$, and that it is Lipschitz from the fact that $\int_{\mathbb{R}^d} \langle \xi \rangle^{1-2s} d\xi$ is finite when $s > (d+1)/2$. We now prove (2.25). Using the Fourier transform definition of the Hilbert norm of $H^{-s}(\mathbb{R}^d)$, we have for any $\rho, \eta \in H^{-s}(\mathbb{R}^d)$, and then for any $\rho, \eta \in \mathbf{P}(E) \subset \mathbf{P}(\mathbb{R}^d) \subset H^{-s}(\mathbb{R}^d)$,

$$\begin{aligned} \|\rho - \eta\|_{H^{-s}}^2 &= \int_{\mathbb{R}^d} (\hat{\rho}(\xi) - \hat{\eta}(\xi)) \overline{(\hat{\rho}(\xi) - \hat{\eta}(\xi))} \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &= \int_{\mathbb{R}^{3d}} (\rho(dx) - \eta(dx)) (\rho(dy) - \eta(dy)) e^{-i(x-y)\xi} \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &= \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes \eta)(dx, dy) + \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\eta^{\otimes 2} - \eta \otimes \rho)(dx, dy). \end{aligned}$$

The last identity follows from (2.25) by choosing $\eta = 0$. \square

Thanks to that Lemma, we will be able to obtain the following key estimate.

Proposition 2.10. *For any $s > \frac{d+1}{2}$ there exists a constant $C = 2\|\Phi_s\|_{Lip} \leq \frac{2^{s+1}c_d}{2s-d-1} \in (0, \infty)$ (where c_d denotes the surface of the unit sphere of \mathbb{R}^d) such that for any $G^N \in \mathbf{P}_{sym}(E^N)$, $N \geq 1$, $f \in \mathbf{P}(E)$, there holds*

$$(2.26) \quad \mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f) \leq C \left[W_1(\hat{G}_2^N, f \otimes f) \right]^{\frac{1}{2}}.$$

PROOF OF PROPOSITION 2.10. Because $\mathbf{P}(E) \subset \mathbf{P}(\mathbb{R}^d) \subset H^{-s}(\mathbb{R}^d)$ for $s > \frac{d}{2}$ and $\Pi(\hat{G}^N, \delta_f) = \{\hat{G}^N \otimes \delta_f\}$, we have

$$\left[\mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f) \right]^2 := \inf_{\pi \in \Pi(\hat{G}^N, \delta_f)} I[\pi] = I(\hat{G}^N \otimes \delta_f),$$

with cost functional

$$I[\pi] := \iint_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho - \eta\|_{H^{-s}}^2 \pi(d\rho, d\eta).$$

Using Lemma 2.9, we have

$$\begin{aligned}
I[\hat{G}^N \otimes \delta_f] &= \int_{\mathbf{P}(E)} \left\{ \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes f)(dx, dy) \right\} \hat{G}^N(d\rho) \\
&\quad + \int_{\mathbf{P}(E)} \left\{ \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (f^{\otimes 2} - f \otimes \rho)(dx, dy) \right\} \hat{G}^N(d\rho) \\
&= \iint_{E^2} \Phi_s(x-y) [\hat{G}_2^N(dx, dy) - \hat{G}_1^N(dx) f(dy)] \\
&\quad + \iint_{E^2} \Phi_s(x-y) [f(dx) f(dy) - f(dx) \hat{G}_1^N(dy)].
\end{aligned}$$

Now we may bound the cost functional as follows:

$$\begin{aligned}
I[\hat{G}^N \otimes \delta_f] &\leq \|\Phi_s\|_{Lip} \left[W_1(\hat{G}_2^N, \hat{G}_1^N \otimes f) + W_1(f \otimes f, f \otimes \hat{G}_1^N) \right] \\
&\leq \|\Phi_s\|_{Lip} \left[W_1(\hat{G}_2^N, f \otimes f) + 2 W_1(f \otimes f, \hat{G}_1^N \otimes f) \right] \\
&\leq \|\Phi_s\|_{Lip} \left[W_1(\hat{G}_2^N, f \otimes f) + W_1(f, \hat{G}_1^N) \right] \\
&\leq 2 \|\Phi_s\|_{Lip} W_1(\hat{G}_2^N, f \otimes f),
\end{aligned}$$

where we have used successively the Katorovich-Rubinstein duality formula (2.3), the triangular inequality, the identity (2.22), and the first inequality in (2.20) together with the fact that $(\hat{G}_2^N)_1 = \hat{G}_1^N$. \square

Putting together Proposition 2.10, Lemma 2.8 above and Lemma 2.3 on comparison of distances in $\mathbf{P}(\mathbf{P}(E))$, we may prove inequality (2.18) of Theorem 2.4.

PROOF OF INEQUALITY (2.18) IN THEOREM 2.4. We define $s := \frac{1}{2\gamma} - \frac{d}{2k}$. Notice that $s > \frac{d+1}{2} \geq 1$ thanks to the conditions satisfied by γ and k . We can thus applied the point i) of Lemma 2.3, Proposition 2.10 and then Lemma 2.8 in order to get

$$\begin{aligned}
\Omega_\infty(G^N; f) &:= W_1(\hat{G}^N, \delta_f) \leq C(d, s) \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f)^{\frac{2k}{d+2ks}} \\
&\leq \frac{C(d, s)}{2s - d - 1} \mathcal{M}_k^{\frac{d}{d+2ks}} W_1(\hat{G}_2^N, f^{\otimes 2})^{\frac{k}{d+2ks}} \\
&\leq \frac{C(d, \gamma, k)}{\gamma^{-1} - d/k - d - 1} \mathcal{M}_k^{\frac{1}{k}} \left(W_1(G_2^N, f^{\otimes 2}) + \frac{2}{N} \right)^\gamma,
\end{aligned}$$

since $\gamma = \frac{k}{d+2ks}$. This is the claimed inequality thanks to the definition of Ω_2 . It is important to notice that the constant $C(d, \gamma, k)$ of the last line depends on d, k and γ via s . But as explained at the end of lemma 2.1, it can be choosen independent of k and γ if $s = \frac{1}{2\gamma} - \frac{d}{2k}$ remains in a compact subset of \mathbb{R}^+ . \square

With stronger moment conditions on the probability measures f and G^N , we may improve the exponent in the right hand side of (2.18) and therefore the rate of convergence to the chaos. Introducing the exponential moment

$$(2.27) \quad \forall F \in \mathbf{P}(E), \quad M_{\beta, \lambda}(F) := \int_E e^{\lambda|x|^\beta} F(dx),$$

$E = \mathbb{R}^d$, $\beta, \lambda > 0$, we have the following result.

Corollary 2.11. (i) *There exists a constant $C = C(d)$ such that if the support of f and G_1^N are both contained in the ball $B(0, R)$, for a positive R , then*

$$(2.28) \quad \Omega_\infty(G^N; f) \leq C R \left(\Omega_2(G^N; f) + \frac{1}{N} \right)^{\frac{1}{d+1}} \left| \ln \left(\Omega_2(G^N; f) + \frac{1}{N} \right) \right|.$$

(ii) *There exists a constant $C = C(d, \beta)$ such that if the f and G_1^N have bounded exponential moment of order $M_{\beta, \lambda}$ for $\beta, \lambda > 0$, there holds*

$$(2.29) \quad \Omega_\infty(G^N; f) \leq \frac{C}{\lambda^{\frac{1}{\beta}}} K^{2(d+1)} \left(\Omega_2(G^N; f) + \frac{1}{N} \right)^{\frac{1}{d+1}} \left| \ln \left(\Omega_2(G^N; f) + \frac{1}{N} \right) \right|^{1+\frac{1}{\beta}}$$

where $K := \max(M_{\beta, \lambda}(f), M_{\beta, \lambda}(G_1^N))$.

PROOF OF COROLLARY 2.11.

Step 1. The compact support case. Here we simply have $M_k(f) \leq R^k$ and the same for the moments of G_1^N . Applying (2.18) with the explicit formula for the constant C , we get for any $0 < \gamma < \frac{1}{d+1}$ and $k > \frac{d}{\gamma^{-1}-d-1}$

$$\Omega_\infty(G^N; f) \leq \frac{C(d, \gamma, k)}{\gamma^{-1} - dk^{-1} - d - 1} R \left(\Omega_2(G^N; f) + \frac{1}{N} \right)^\gamma.$$

And we use the remark at the end of the previous proof that allows to replace $C(d, \gamma, k)$ by $C(d)$ if $s = \frac{1}{2\gamma} - \frac{d}{2k}$ is restricted to some compact subspace of $[1, +\infty)$. It will be the case in the sequel since we shall choose k large and γ close to $\frac{1}{d+1}$. Letting $k \rightarrow +\infty$ leads to

$$\Omega_\infty(G^N; f) \leq \frac{C(d)}{\gamma^{-1} - d - 1} R \left(\Omega_2(G^N; f) + \frac{1}{N} \right)^\gamma.$$

Denoting $\alpha := \frac{1}{\gamma} - d - 1$ and $a = \Omega_2(G^N; f) + \frac{1}{N}$ which we assume smaller than $\frac{1}{2}$, the r.h.s can be rewritten

$$\Omega_\infty(G^N; f) \leq C(d) \frac{R}{\alpha} a^{1/(d+1+\alpha)}.$$

Some optimization leads to the natural choice $\alpha = 2 \frac{(d+1)^2}{|\ln a|}$. It comes

$$\Omega_\infty(G^N; f) \leq C(d) R |\ln a| a^{1/(d+1)} a^{1/(d+1+\alpha)-1/(d+1)}.$$

Since $\frac{1}{d+1} - \frac{1}{d+1+\alpha} \leq \frac{\alpha}{(d+1)^2} \leq \frac{1}{2|\ln a|}$, we deduce

$$a^{1/(d+1+\alpha)-1/(d+1)} \leq a^{-1/(2|\ln a|)} = e^{\frac{1}{2}}$$

and this concludes the proof of point (i).

Step 2. The case of exponential moment.

Using the elementary inequality $x^k \leq \left(\frac{k}{\lambda\beta e}\right)^{k/\beta} e^{\lambda|x|^\beta}$, we get the following bound on the k moment

$$M_k(F)^{1/k} \leq \left(\frac{k}{\lambda\beta e}\right)^{1/\beta} M_{\beta, \lambda}(F)^{1/k},$$

and it implies with our notations $\mathcal{M}_k^{\frac{1}{k}} \leq \left(\frac{k}{\lambda\beta e}\right)^{1/\beta} (2K)^{1/k}$. Applying (2.18) with the explicit formula for the constant C and the notation a of the previous step, we get for any $0 < \gamma < \frac{1}{d+1}$ and $k > \frac{d}{\gamma^{-1}-d-1}$

$$\Omega_\infty(G^N; f) \leq \frac{C(d)}{(\lambda\beta e)^{1/\beta}} \frac{k^{1/\beta}}{\gamma^{-1} - dk^{-1} - d - 1} K^{1/k} a^\gamma.$$

Here we cannot take the limit as $k \rightarrow \infty$, but optimizing in k the second fraction of the r.h.s, we choose k satisfying $\frac{1}{\gamma} - d - 1 = \frac{2d}{k}$ and get the bound

$$\Omega_\infty(G^N; f) \leq \frac{C(d, \beta)}{\lambda^{1/\beta}} \frac{4d}{(\gamma^{-1} - d - 1)^{1+1/\beta}} K^{1/k} a^\gamma.$$

Still denoting $\alpha = \frac{1}{\gamma} - d - 1 = \frac{2d}{k}$, the choice $\alpha = 2 \frac{(d+1)^2}{|\ln a|}$ leads this time to the bound

$$\Omega_\infty(G^N; f) \leq \frac{C(d, \beta)}{\lambda^{1/\beta}} K^{(d+1)/\ln 2} |\ln a|^{1+1/\beta} a^{1/(d+1)},$$

which concludes the proof. \square

Remark 2.12. *Inequality (2.18) in Theorem 2.4 says in particular that for any $k > 0$ and $0 < \gamma < (d + 1 + d/k)^{-1}$ there exists a constant $C := C(d, \gamma, k)$ such that for any $f \in \mathbf{P}(E)$, there holds*

$$(2.30) \quad \Omega_\infty(f^{\otimes N}; f) \leq \frac{C M_k(f)^{1/k}}{N^\gamma}.$$

For such a tensor product probability measures framework, the above rate can be improved in the following way.

Theorem 2.13 ([55, 10]). *1. For a moment weight exponent $k > 0$ and an exponent*

- (i) $\gamma = \gamma_c := (2 + 1/k)^{-1}$ when $d = 1$,
- (ii) $\gamma \in (0, \gamma_c)$ with $\gamma_c := (2 + 2/k)^{-1}$ when $d = 2$,
- (iii) $\gamma = \gamma_c := (d + d/k)^{-1}$ when $d \geq 3$,

there exists a finite constant $C := C(d, \gamma, k)$ such that (2.30) holds.

2. Moreover, for any moment weight exponents $\lambda, \beta > 0$, there exists a finite constant $C := C(d, \lambda, \beta, M_{\beta, \lambda}(f))$ such that

$$(2.31) \quad \Omega_\infty(f^{\otimes N}; f) \leq C \frac{(\ln N)^{1/\beta}}{N^{1/2}}, \text{ if } d = 1, \quad \Omega_\infty(f^{\otimes N}; f) \leq C \frac{(\ln N)^{1+1/\beta}}{N^{1/d}}, \text{ if } d \geq 2.$$

On the one hand, using similar Hilbert norm arguments as those used in the proof of Proposition 2.10 and inequality (2.18) in Theorem 2.4, the first point in Theorem 2.13 has been proved in [55, Lemma 4.2(iii)] with however the restriction $\gamma \in (0, \gamma_c)$ when $d \geq 1$. The optimal rate $\mathcal{O}(1/N^{(2+1/k)^{-1}})$ in the critical case $\gamma = \gamma_c$, $d = 1$, is not mentioned in [55, Lemma 4.2(iii)] but follows from a careful but straightforward reading of the proof of [55, Lemma 4.2(iii)]. The better rate obtained in Theorem 2.13 with respect to (2.30) is due to the fact that for a tensor product measure one can work in the Hilbert space H^{-s} with $s > d/2$ rather than with $s > (d + 1)/2$ in the general case. The second point in Theorem 2.13 follows by adapting the proof of Corollary 2.11 to this tensor product measures framework.

On the other hand, using matching techniques, it has been proved in [26, 10] that (2.30) also holds true for the critical exponent $\gamma_c = 1/d$ in the compact support case (or exponential moment with $\beta = 1$) when $d \geq 3$ and $\gamma_c = (d + d/k)^{-1}$ in the case of finite moment of order k when $d \geq 3$. These last results thus slightly improve the estimates available thanks to our Hilbert norms technique. It is worth mentioning that the critical exponents are known to be optimal, see for instance [26, 4]. A natural question is whether the rates in inequality (2.18) and in Corollary 2.11 may be improved using similar arguments as in [26, 10].

We come to the proof of the last part of Theorem 2.4, which will be a consequence of the following proposition

Proposition 2.14. *For $F^N, G^N \in \mathbf{P}_{\text{sym}}(E^N)$, there holds*

$$(2.32) \quad W_1(F^N, G^N) = \mathcal{W}_1(\hat{F}^N, \hat{G}^N).$$

PROOF OF PROPOSITION 2.14. We split the proof into two steps.

Step 1. A reformulation of the problem. Since we are dealing with symmetric probability measures, it is natural to introduce the equivalence relation \sim in E^N by saying that $X = (x_1, \dots, x_N), Y = (y_1, \dots, y_N) \in E^N$ are equivalent, we write $X \sim Y$, if there exists a permutation $\sigma \in \mathfrak{S}_N$ such that $Y = X_\sigma := (x_{\sigma(1)}, \dots, x_{\sigma(N)})$.

We also introduce on E^N the "semi"-distance w_1

$$(2.33) \quad w_1(X, Y) := \inf_{\sigma \in \mathfrak{S}_N} d_{E^N}(X, Y_\sigma) = \inf_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_{\sigma(i)}),$$

which only satisfies $w_1(X, Y) = 0$ iff $X \sim Y$. We then introduce the associated MKW functionnal W_1^\dagger . For $F^N, G^N \in \mathbf{P}_{\text{sym}}(E^N)$,

$$W_1^\dagger(F^N, G^N) := \inf_{\pi^N \in \Pi(F^N, G^N)} \int_{E^N \times E^N} w_1(X, Y) \pi^N(dX, dY).$$

It is in fact a distance on the space of symmetric probability measures, but this point will also be a consequence of our proof. It is a classical result (see for instance [72, Introduction. Example: the discrete case]) that

$$(2.34) \quad \forall X, Y \in E^N, \quad W_1(\mu_X^N, \mu_Y^N) = w_1(X, Y),$$

(shortly, it means than we do not need to split the small Dirac masses when we try to optimize the transport between two empirical measures). We recall the notation p_N defined in section 2.1 for the application that sends a configuration to the associated empirical measure : $p_N(X) = \mu_X^N$.

Remark that its associated push-forward mapping restricted to the symmetric probability measures

$$\tilde{p}_N : \mathbf{P}_{\text{sym}}(E^N) \rightarrow \mathbf{P}(\mathcal{P}_N(E)) \subset \mathbf{P}(\mathbf{P}(E)), \quad G^N \mapsto \hat{G}^N := G_{\#}^N p_N,$$

is a bijection. Its inverse can be simply expressed thanks to a dual formulation: for $\alpha \in \mathbf{P}(\mathcal{P}_N(E))$, its inverse $\tilde{\alpha} = \tilde{p}_N^{-1} \alpha$ is the probability measure satisfying

$$\forall \varphi \in C_b(E^N), \quad \int_{E^N} \varphi(X) \tilde{\alpha}(dX) = \int_{\mathcal{P}_N(E)} \tilde{\varphi}(\rho) \alpha(d\rho),$$

where $\tilde{\varphi}(\rho) := \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi(X_\sigma)$, for any given X such that $\mu = \mu_X^N$. Similarly, defining $\mathbf{P}_{s,s}(E^N \times E^N)$ the subset of $\mathbf{P}(E^N \times E^N)$ of probability measures which are invariant under permutations on the first and second blocks of N variables separately, we have that

$$\tilde{p}_N^{\otimes 2} : \mathbf{P}_{s,s}(E^N \times E^N) \rightarrow \mathbf{P}(\mathcal{P}_N(E) \times \mathcal{P}_N(E)), \quad \pi^N \mapsto \hat{\pi}^N := \pi_{\#}^N(p_N, p_N),$$

is a bijection.

The identity (2.34) and the bijection \tilde{p}_N allows us to establish the identity

$$(2.35) \quad \forall F^N, G^N \in \mathbf{P}(E^N), \quad W^\dagger(F^N, G^N) = \mathcal{W}_1(\hat{F}^N, \hat{G}^N).$$

Indeed, denoting $\Pi_{s,s}(F^N, G^N) = \Pi(F^N, G^N) \cap \mathbf{P}_{s,s}(E^N, E^N)$, we have

$$\begin{aligned}
W_1^\dagger(F^N, G^N) &= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{E^N \times E^N} w_1(X, Y) \pi^N(dX, dY) \\
&= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{E^N \times E^N} W_1(p_N(X), p_N(Y)) \pi^N(dX, dY) \\
&= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{\mathcal{P}_N(E) \times \mathcal{P}_N(E)} W_1(\rho, \eta) \pi_\#^N(p_N, p_N)(d\rho, d\eta) \\
&= \inf_{\hat{\pi} \in \Pi(\hat{F}^N, \hat{G}^N)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho, \eta) \hat{\pi}(d\rho, d\eta) = \mathcal{W}_1(\hat{F}^N, \hat{G}^N),
\end{aligned}$$

where we have essentially used the invariance $w_1(X, Y) = w_1(X_\sigma, Y_\tau)$ for any $\sigma, \tau \in \mathfrak{S}_N$ and the fact that $\hat{p}_N^{\otimes 2}$ is a bijection.

Step 2. The equality $W_1^\dagger = W_1$. The interest of the reformulation (2.35) is that we can now work on one space: E^N . Remark that since $w_1(X, Y) \leq d_{E^N}(X, Y)$, we always have $W_1^\dagger \leq W_1$, and the equality will hold only if one transference plan for W_1^\dagger is concentrated on the set

$$\mathcal{C} := \left\{ (X, Y) \in E^N \times E^N \text{ s.t. } w_1(X, Y) = \inf_{\sigma \in \mathfrak{S}_N} d_{E^N}(X, Y_\sigma) = d_{E^N}(X, Y) \right\}.$$

We choose an optimal transference plan π for W_1^\dagger . For simplicity we will assume that π is symmetric, i.e. unchanged by the applications $P_\sigma : (X, Y) \mapsto (X_\sigma, Y_\sigma)$ for any $\sigma \in \mathfrak{S}_N$. If not, we replace it by its symmetrization $\frac{1}{N!} \sum_{\sigma} \pi_\# P_\sigma$ which will still be an optimal transference plan of F^N onto G^N . Starting from π , we will construct a transference plan $\pi^* \in \Pi(F^N, G^N)$ such that

- i) π^* is concentrated on \mathcal{C} .
- ii) $I_N[\pi] = \int w_1(X, Y) \pi(dX, dY) = \int w_1(X, Y) \pi^*(dX, dY) = I_N[\pi^*]$

Both properties imply then that

$$\begin{aligned}
W_1^\dagger(F^N, G^N) &= \int_{E^N \times E^N} w_1(X, Y) \pi(dX, dY) = \int_{E^N \times E^N} w_1(X, Y) \pi^*(dX, dY) \\
&= \int_{E^N \times E^N} d_{E^N}(X, Y) \pi^*(dX, dY) \geq W_1(F^N, G^N)
\end{aligned}$$

which is the desired inequality.

We define π^* in the following way. First, we introduce for any $X, Y \in E^N$

$$\begin{aligned}
\mathcal{C}_{X;Y} &:= \{Z \in E^N; Z \sim Y \text{ and } d_{E^N}(X, Z) = w_1(X, Y)\} \subset E^N \\
\rho_{X;Y} &:= \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \delta_{(X,Z)} \in \mathbf{P}(E^N \times E^N), \quad N_{X;Y} := \#\mathcal{C}_{X;Y} \in \mathbb{N}^*.
\end{aligned}$$

We note that $Z \in \mathcal{C}_{X;Y}$ iff $Z \sim Y$ and $(X, Z) \in \mathcal{C}$, so that $\text{Supp } \rho_{X;Y} \subset \mathcal{C}$. It can be shown that $(X, Y) \mapsto N_{X;Y}$ is a borelian application (it takes finite values and its level set are closed) and that $E^N \times E^N \rightarrow \mathbf{P}(E^N \times E^N)$, $(X, Y) \mapsto \rho_{X;Y}$ is also borelian if $\mathbf{P}(E^N \times E^N)$ is endowed with the weak topology of measures. This allows us to define a transference plan π^* by

$$\pi^* := \int_{E^N \times E^N} \rho_{X;Y} \pi(dX, dY) \in \mathbf{P}(E^N \times E^N),$$

or in other words, for any $\psi \in C_b(E^N \times E^N)$, we have

$$\begin{aligned} \langle \pi^*, \psi \rangle &= \int_{E^{2N}} \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \int_{E^{2N}} \psi(X', Y') \delta_{(X,Z)}(dX', dY') \pi^N(dX, dY) \\ &= \int_{E^{2N}} \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \psi(X, Z) \pi^N(dX, dY). \end{aligned}$$

It remains to proof that π^* satisfy the announced properties. Since $\rho_{X;Y}$ is supported in \mathcal{C} for any $(X, Y) \in E^N \times E^N$, it is also the case for π^* . It is also not difficult to show that the transport cost for w_1 is preserved. Indeed, we have

$$\begin{aligned} \int_{E^{2N}} d_{E^N}(X', Y') \pi^*(dX', dY') &= \int_{E^{2N}} \left(\frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} d_{E^N}(X, Z) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \left(\frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} w_1(X, Y) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} w_1(X, Y) \pi(dX, dY). \end{aligned}$$

The fact that π^* has first marginal F^N is also clear since for any $\varphi \in C_b(E^N)$

$$\begin{aligned} \int_{E^{2N}} \varphi(X') \pi^*(dX', dY') &= \int_{E^{2N}} \left(\frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \varphi(X) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \varphi(X) \pi(dX, dY) = \int_{E^N} \varphi(X) F^N(dX). \end{aligned}$$

For the second marginal, we shall use the following properties of $\mathcal{C}_{X;Y}$ and $N_{X;Y}$

$$\forall \tau \in \mathfrak{S}_N, \quad Z_\tau \in \mathcal{C}_{X_\tau; Y_\tau} \Leftrightarrow Z \in \mathcal{C}_{X; Y}, \quad \text{and thus} \quad N_{X_\tau; Y_\tau} = N_{X; Y}.$$

Thanks to the invariance by symmetry of π and G^N , we can write for any $\varphi \in C_b(E^N)$

$$\begin{aligned} \int_{E^{2N}} \varphi(Y) \pi^*(dX, dY) &= \int_{E^{2N}} \left(\frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \varphi(Z) \right) \pi(dX, dY) \\ &= \frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \int_{E^{2N}} \left(\frac{1}{N_{X_\tau; Y_\tau}} \sum_{Z \in \mathcal{C}_{X_\tau; Y_\tau}} \varphi(Z) \right) \pi(dX, dY) \\ &= \frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \int_{E^{2N}} \left(\frac{1}{N_{X; Y}} \sum_{Z \in \mathcal{C}_{X; Y}} \varphi(Z_\tau) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \left(\frac{1}{N_{X; Y}} \sum_{Z \in \mathcal{C}_{X; Y}} \tilde{\varphi}(Z) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \tilde{\varphi}(Y) \pi(dX, dY) \\ &= \int_{E^{2N}} \tilde{\varphi}(Y) G^N(dX) = \int_{E^{2N}} \varphi(Y) G^N(dX), \end{aligned}$$

where we have introduced the symmetrization of φ defined by $\tilde{\varphi}(Z) := \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi(Z_\sigma)$ and we have used that $\tilde{\varphi}(Z) = \tilde{\varphi}(Y)$ for any $Z \in \mathcal{C}_{X;Y}$ and the fact that G^N is symmetric. This concludes the proof. \square

Putting together Proposition 2.14 and (2.30), we obtain the inequality (2.19) of Theorem 2.4.

PROOF OF INEQUALITY (2.19) IN 2.4. We have

$$\begin{aligned} |\Omega_N(G^N, f) - \Omega_\infty(G^N, f)| &= |W_1(G^N, f^{\otimes N}) - W_1(\hat{G}^N, \delta_f)| \\ &= |W_1(\hat{G}^N, \widehat{f^{\otimes N}}) - W_1(\hat{G}^N, \delta_f)| \\ &\leq W_1(\widehat{f^{\otimes N}}, \delta_f) = \Omega_\infty(f^{\otimes N}; f) \\ &\leq \frac{C M_k(f)^{1/k}}{N^\gamma}, \end{aligned}$$

where we have used the definition of Ω_N , Ω_∞ , the triangular inequality, Proposition 2.14 and (2.30). \square

3. ENTROPY CHAOS AND FISHER INFORMATION CHAOS

In this section $E \subset \mathbb{R}^d$ stands for an open set or the adherence of a open space (so that the gradient of a function on E is well defined).

3.1. Entropy chaos. The entropy of a probability measure on a compact subset of \mathbb{R}^d with density $f dx$ is well defined by the formula $\int f \ln f$. On a (possibly) unbounded set E , we have to be more careful because the entropy may not be defined for probability measure decreasing too slowly at infinity. This is a well known issue, but we present here a rigorous definition for probability measures $F \in \mathbf{P}(E^j)$ having a finite moment M_k for some $k > 0$. It will be useful in the section 5 where we define the level 3 entropy and Fisher information on $\mathbf{P}(\mathbf{P}(E))$.

We emphasize that in the sequel we shall use the same notation F for a probability measure and its density $F dx$ with respect to the Lebesgue measure, when the last quantity exists. For any $k > 0$ and $F \in \mathbf{P}_k(E^j) \cap L^1$, we define the (opposite of the Boltzmann's) entropy

$$\begin{aligned} (3.1) \quad H_j(F) &:= \int_{E^j} F \log F \\ &= \int_E h(F/G_k^j) G_k^j + \int_E F \log G_k^j \quad (= H_j^{(1)}(F)) \end{aligned}$$

with $G_k^j(V) := c_k^j \exp(-|v_1|^k - \dots - |v_j|^k) \in \mathbf{P}(E^j)$, c_k chosen so that G_k^j is a probability measure, and $h(s) := s \log s - s + 1$. The RHS term is well defined in $\mathbb{R} \cup \{+\infty\}$ as the sum of a nonnegative term and a finite real number, and it can be checked that it is equal to the middle term, which has thus a sense. Next, we extend the entropy functional to any $F \in \mathbf{P}_k(E^j)$ by setting

$$(3.2) \quad H_j(F) := \sup_{\phi_j \in C_b(E^j)} \left\{ \langle F, \phi_j \rangle - H^*(\phi_j) \right\} + \int_E F \log G_k^j \quad (= H^{(2)}(F))$$

where

$$H^*(\phi_j) := \int_{E^j} h^*(\phi_j) G_k^j$$

and where $h^*(t) := e^t - 1$ is the Legendre transform of h . Finally, we define the normalized entropy functional H by

$$(3.3) \quad \forall F \in \mathbf{P}_k(E^j) \quad H(F) := \frac{1}{j} H_j(F).$$

We start recalling without proof a very classical result concerning the entropy.

Lemma 3.1. *Let us fix $k > 0$. The entropy functional $\mathbf{P}_k(E) \rightarrow \mathbb{R} \cup \{+\infty\}$, $\rho \mapsto H_j(\rho)$ is well defined by the expression (3.2), is convex and is l.s.c. for the following notion of converging sequences: $\rho_n \rightarrow \rho$ in the weak sense of measures in $\mathbf{P}(E)$ and $\langle \rho_n, |v|^m \rangle$ is bounded for some $m > k$ (the same holds of course for H). Moreover, $H_j(F)$ does not depend on the choice of k used in the expression (3.2),*

$$H(F) \geq \log c_k - M_k(F) \quad \forall F \in \mathbf{P}_k(E),$$

and $H(F) < \infty$ iff $F \in L^1$, $F \log F \in L^1(E)$, and then $H(F) = H^{(1)}(F)$.

We also recall the definition of the (non-normalized) relative entropy between two probability measures ρ and η of $\mathbf{P}(E^j)$:

$$(3.4) \quad H_j(\rho|\eta) := \int_{E^j} \ln \left(\frac{d\rho}{d\eta} \right) d\rho = \int_{E^j} (g \ln g + 1 - g) d\eta$$

with $g = \frac{d\rho}{d\eta}$ if ρ is absolutely continuous with respect to η . If g is not defined, then $H_j(\rho|\eta) := +\infty$. The associated normalized quantity is simply $H(\rho|\eta) := \frac{1}{j} H_j(\rho|\eta)$. The relative entropy is defined without moment assumption since the quantity under the last integral is nonnegative. It can also be defined using a dual formula similar to (3.2). For a fixed η it has the same properties as the entropy.

We now give two elementary and well known results which are fundamental for the analysis of the entropy defined on space product.

Lemma 3.2. *On $\mathbf{P}_m(E^j)$, $m > 0$, the entropy satisfies the identity*

$$(3.5) \quad \forall f \in \mathbf{P}_m(E) \quad H(f^{\otimes j}) = H(f).$$

PROOF OF LEMMA 3.2. If $f \in \mathbf{P}_m(E)$ is a function such that $H(f) < \infty$, then we may use (3.1) as a definition and

$$H(f^{\otimes j}) = \frac{1}{j} \int_{E^j} f^{\otimes j} \log f^{\otimes j} = \int_{E^j} f^{\otimes j}(v_1, \dots, v_j) \log f(v_1) = H_1(f).$$

In the contrary, $H_1(f) = \infty$ implies $H_j(f^{\otimes j}) = \infty$. □

Lemma 3.3. (i) *For any functions $f, g \in L_m^1(E) \cap \mathbf{P}(E)$, $m > 0$, there holds*

$$(3.6) \quad H(f) := \int_E f \log f \geq \int_E f \log g, \quad \text{or} \quad H(f|g) := \int_E f \log(f/g) \geq 0,$$

with equality only if $f = g$ a.e..

(ii) *More generally, for any nonnegative functions $f, g \in L_m^1(E)$, $m > 0$, there holds*

$$\int_E f \log \frac{f}{g} \geq F \log \frac{F}{G}, \quad \text{with} \quad F := \int_E f, \quad G := \int_E g.$$

(iii) *A consequence of (i) is that if $F \in \mathbf{P}(E^j)$ has first marginal f with $H(f) < +\infty$, then*

$$H(F) \geq H(f) \quad \text{with equality only if } F = f^{\otimes j} \text{ a.e..}$$

(iv) The entropy is superadditive: for any $F \in \mathbf{P}_m(E^{i+j}) \cap \mathbf{P}_{\text{sym}}(E^{i+j})$, $i, j \in \mathbb{N}^*$, $m > 0$, the following inequality holds

$$(3.7) \quad H_{i+j}(F_{i+j}) \geq H_i(F_i) + H_j(F_j), \quad (\text{non-normalized entropy}),$$

where F_ℓ as usual stands for the ℓ -th marginal of F .

PROOF OF LEMMA 3.3. (i) To obtain the inequality, write $H(f|g) = \int h(f/g)f$ and use the fact that $h(s) = s \log s - s + 1$ is a nonnegative function. Next there is equality only if $h(f/g) = 0$ a.e. on $\{f > 0\}$. Since h vanishes only at $s = 1$, it means that $f = g$ a.e. on $\{f > 0\}$. Using that $\int f = \int g = 1$, we obtain the claimed equality.

(ii) We write

$$\int_E f \log \frac{f}{g} = F \int_E f/F \log \frac{f/F}{g/G} + \int_E f \log \frac{F}{G},$$

the first term is nonnegative thanks to (3.6) and the second term is the one which appears on the RHS of the claimed inequality.

(iii) We use the first inequality (3.6) on E^j with F and $f^{\otimes j}$

$$H(F) = \frac{1}{j} \int_{E^j} F \log F \geq \frac{1}{j} \int_{E^j} F \log f^{\otimes j} = \int_{E^j} F(V) \log f(v_1) dV = H(f).$$

Using again the point i), we see that equality can occur only if $F = f^{\otimes j}$ a.e..

(iv) Denote $h_\ell := H_\ell(F_\ell)$. If $h_{i+j} = +\infty$ there is nothing to prove. Otherwise, we have $h_{i+j} < \infty$ which in turn implies $F \in L^1(E^{i+j})$, then $F_i \in L^1(E^i)$, $F_j \in L^1(E^j)$, so that the entropy may be defined thanks to (3.1). In $\mathbb{R} \cup \{-\infty\}$, we compute

$$\begin{aligned} h_{i+j} - h_i - h_j &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} \\ &\quad - \int_{E^{i+j}} F_{i+j} \log F_i(v_1, \dots, v_i) - \int_{E^{i+j}} F_{i+j} \log F_j(v_{i+1}, \dots, v_{i+j}) \\ &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} - \int_{E^{i+j}} F_{i+j} \log F_i \otimes F_j \geq 0, \end{aligned}$$

thanks to (3.6). □

Our first result shows that entropy chaos is a stronger notion than Kac's chaos.

Theorem 3.4 (Entropy and chaos). *Consider (G^N) a sequence of $\mathbf{P}_{\text{sym}}(E^N)$ such that $\langle G_1^N, |v|^m \rangle \leq a$ for any $N \geq 1$ and for some fixed $m, a > 0$ and consider $f \in \mathbf{P}(E)$.*

1) *If $G_j^N \rightharpoonup F_j$ weakly in $\mathbf{P}(E^j)$ for some given $j \geq 1$, then*

$$(3.8) \quad H(F_j) \leq \liminf H(G^N).$$

In particular, when (G^N) is f -Kac's chaotic, (3.8) holds for any $j \geq 1$ with $F_j := f^{\otimes j}$.

2) *On the other way round, if (G^N) is f -entropy chaotic, then (G^N) is f -Kac's chaotic.*

PROOF OF THEOREM 3.4. *Step 1.* For any $N \geq j$ we introduce the Euclidean decomposition $N = nj + r$, $0 \leq r \leq j - 1$, exactly as in the proof of Proposition 2.6. Iterating n times the superadditivity inequality (3.7) we have

$$H_N(F^N) \geq n H_j(F_j^N) + H(F_r^N),$$

with the convention $H(F_r^N) = 0$ when $r = 0$. We get (3.8) by passing to the limit in that inequality divided by N , using that H is l.s.c. and that $H(F_r^N)$ is bounded by below thanks to Lemma 3.1 and the condition on the moment.

Step 2. We assume that (G^N) is f -entropy chaotic, that is

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad H(G^N) \rightarrow H(f) < \infty.$$

Let us fix $j \geq 1$. The sequence (G_j^N) being bounded in $\mathbf{P}_m(E^j)$, there exists $F_j \in \mathbf{P}(E^j)$ and a subsequence $(G_j^{N'})$ such that $G_j^{N'} \rightharpoonup F_j$ weakly in $\mathbf{P}(E^j)$. Thanks to step 1, we have

$$H(F_j) \leq \liminf H(G_j^{N'}) \leq \liminf H(G^{N'}) = H(f) = H(f^{\otimes j}).$$

Since the first marginal of F_j is $(F_j)_1 = \lim_{N \rightarrow +\infty} G_1^N = f$, the third point of Lemma 3.3 gives that $F_j = f^{\otimes j}$ a.e.. As a conclusion and because we have identified the limit, we have proved that the all sequence (G_j^N) weakly converges to $f^{\otimes j}$. \square

3.2. Fisher chaos. We now establish similar results for the Fisher information functional. For an arbitrary probability measure $G \in \mathbf{P}(E^j)$, we define the normalized Fisher information by

$$(3.9) \quad I_j^{(1)}(G) := \begin{cases} \int_{E^j} \frac{|\nabla G|^2}{G} = \int_{E^j} |\nabla \ln G|^2 G & \text{if } G \in W^{1,1}(E^j), \\ +\infty & \text{if } G \notin W^{1,1}(E^j), \end{cases}$$

For $G \in \mathbf{P}(E^j)$, we also give an alternative definition

$$(3.10) \quad I_j^{(2)}(G) := \sup_{\psi \in C_b^1(E^j)^d} \langle G, -\frac{|\psi|^2}{4} - \operatorname{div} \psi \rangle \in \mathbb{R} \cup \{+\infty\}.$$

Lemma 3.5. *For all $j \in \mathbb{N}$, the identity $I_j^{(1)} = I_j^{(2)}$ holds on $\mathbf{P}(E^j)$, and we simply denoted by I_j the usual (non-normalized) Fisher information and by $I = j^{-1} I_j$ the normalized Fisher information. The functionals I_j and I are proper, convex, l.s.c. (in the sense of the weak convergence of measures) on $\mathbf{P}(E^j)$.*

PROOF OF LEMMA 3.5. For the sake of simplicity, we only deal with the case $j = 1$. We split the proof into two steps.

Step 1. Assume that $f \in W^{1,1}$. Since for all $\psi \in C_b^1(E)^d$

$$|\nabla \ln f|^2 - \nabla \ln f \cdot \psi + \frac{|\psi|^2}{4} = \left| \nabla \ln f - \frac{\psi}{2} \right|^2 \geq 0,$$

we have

$$I^{(1)}(f) = \int_E |\nabla \ln f|^2 f \geq \int_E \left(\nabla \ln f \cdot \psi - \frac{|\psi|^2}{4} \right) f.$$

For any sequence (ψ_n) of smooth functions approximating $2\nabla \ln f = 2\frac{\nabla f}{f}$, we obtain that

$$(3.11) \quad \begin{aligned} I^{(1)}(f) &= \lim_{n \rightarrow \infty} \int_E \left(\nabla \ln f \cdot \psi_n - \frac{|\psi_n|^2}{4} \right) f \\ &= \sup_{\psi \in C_b^1(E)^d} \int_E \left(\nabla \ln f \cdot \psi - \frac{|\psi|^2}{4} \right) f \\ &= \sup_{\psi \in C_b^1(E)^d} \int_E \left[\nabla f \cdot \psi - f \frac{|\psi|^2}{4} \right] =: I^{(3)}(f). \end{aligned}$$

The remaining equality $I^{(3)} = I^{(2)}$ is just a simple integration by parts. Remark that maximizing sequences (ψ_n) must converge (up to some subsequence) pointwise to $2\nabla \ln f$ a.e. on $\{f \neq 0\}$. We shall use that point in the sequel.

We also remark that this reformulation $I^{(2)}$ is also exactly the one obtained when using the general Fenchel-Moreau theorem on the convex function $(a, b) \rightarrow \frac{|b|^2}{a}$ (which is used in the integral defining $I^{(1)}$).

Step 2. It remains to check that the equality $I^{(1)} = I^{(2)}$ is also true on $\mathbf{P}(E) \setminus W^{1,1}(E)$. In other words that if $f \notin W^{1,1}(E)$ then $I^{(2)}(f) = +\infty$. In what follows, we prove the contraposition : $I^{(2)}(f) < +\infty$ implies $f \in W^{1,1}(E)$. Once it will be done, we will have $I^{(1)} = I^{(2)}$ everywhere, from what follows that I is l.s.c. in the sense of the weak convergence of measures.

Consider $f \in \mathbf{P}(E)$ and assume $I^{(2)}(f) < \infty$. We deduce that for any $\psi \in C_b^1(E)^d$ and any $t \in \mathbb{R}$

$$\int_E f \left[-t^2 \frac{|\psi|^2}{4} - t \operatorname{div} \psi \right] \leq I^{(2)}(f),$$

so that by optimizing in $t \in \mathbb{R}$ and using that $f \in \mathbf{P}(E)$, we get

$$\forall \psi \in C_b^1(E)^d \quad \left| \int_E f \operatorname{div} \psi \right|^2 \leq 4 I^{(2)}(f) \int_E f \frac{|\psi|^2}{4} \leq I^{(2)}(f) \|\psi\|_{L^\infty}^2.$$

That inequality implies $f \in BV(E)$ and $\|\nabla f\|_{TV} \leq \sqrt{I^{(2)}(f)}$. Using that $f \in BV(E)$ and making an integration by part in the definition of $I^{(2)}(f)$, we find

$$I^{(2)}(f) = \sup_{\psi \in C_b^1(E)^d} \int_E [\nabla f \cdot \psi - f \frac{|\psi|^2}{4}] = I^{(3)}(f).$$

Now, for any compact subset $K \subset E$ with zero Lebesgue measure, we may find a sequence $\rho_\varepsilon \in C_c^1(E)$ such that $0 \leq \rho_\varepsilon \leq 1$, $\rho_\varepsilon = 1$ on K and $\rho_\varepsilon \rightarrow 0$ a.e., so that for any $t > 0$ and using that $f \in BV(E) \subset L^1(E)$, we get for all $\varepsilon > 0$

$$\begin{aligned} t \int_K |\nabla f| &\leq t \int_E |\nabla f| \rho_\varepsilon \\ &\leq \sup_{\psi \in C_c^1(E)^d, \|\psi\|_\infty \leq 1} \int_E \nabla f \cdot \psi t \rho_\varepsilon \\ &\leq \sup_{\psi \in C_c^1(E)^d, \|\psi\|_\infty \leq 1} \int_E \left[\nabla f \cdot t \psi \rho_\varepsilon - f t^2 \frac{|\psi|^2 \rho_\varepsilon^2}{4} \right] + \frac{t^2}{4} \int_E f \rho_\varepsilon^2 \\ &\leq I^{(3)}(f) + \frac{t^2}{4} \int_E f \rho_\varepsilon^2. \end{aligned}$$

Passing to the limit $\varepsilon \rightarrow 0$ using that $f \in L^1(K)$ and then $t \rightarrow \infty$, we deduce that ∇f vanishes on K , which precisely means that ∇f is a measurable function. We have proved $f \in W^{1,1}(\mathbb{R}^d)$. \square

Similarly, we define for two measures ρ and η on E^j their (non-normalized) relative Fisher information $I(\rho|\eta)$ by

$$(3.12) \quad I_j(\rho|\eta) := \int_{E^j} \frac{|\nabla g|^2}{g} d\eta = \int_{E^j} \left| \nabla \ln \frac{d\rho}{d\eta} \right|^2 d\rho,$$

where $g = \frac{d\rho}{d\eta}$ if ρ is absolutely continuous with respect to η . If not, $I_j(\rho|\eta) := +\infty$. The associated normalized quantity is simply $I(\rho|\eta) := \frac{1}{j} I_j(\rho|\eta)$. For a fixed η , the

relative Fisher information has roughly the same properties as the Fisher information. In particular, if η has a derivable density, we have the equality

$$(3.13) \quad I_j(\rho|\eta) = \sup_{\varphi \in C_b^1(E^j)^{dj}} \int_{\mathbb{R}^j} \left(-\varphi \cdot \frac{\nabla \eta}{\eta} - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) d\rho.$$

Lemma 3.6. *For any $f \in \mathbf{P}(E)$ there holds $I(f^{\otimes j}) = I(f)$.*

PROOF OF LEMMA 3.6. If $I(f) < \infty$ then $f \in W^{1,1}(E)$ and also $f^{\otimes j} \in W^{1,1}(E^j)$. The following computation is then meaningful

$$I(f^{\otimes j}) = \frac{1}{j} \int_{E^j} \frac{|\nabla_{E^j} f^{\otimes j}|^2}{f^{\otimes j}} = \int_{E^j} \frac{|\nabla_E f|^2}{f} \otimes f^{\otimes(j-1)} = I(f).$$

Since $I_j(f^{\otimes j}) < \infty$ implies $f^{\otimes j} \in W^{1,1}(E^j)$ and then $f \in W^{1,1}(E)$, we also have $I_j(f^{\otimes j}) = j I(f)$ if $I(f) = \infty$. \square

Lemma 3.7. *For any $F \in P_{\text{sym}}(E^j)$ and $1 \leq \ell \leq j$, then holds*

- (i) $I(F_\ell) \leq I(F)$.
- (ii) *The Fisher information is super-additive. It means that*

$$(3.14) \quad I_j(F) \geq I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell}), \quad (\text{non-normalized Fisher information}),$$

with in the case $I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell}) < +\infty$ equality only if $F = F_\ell \otimes F_{j-\ell}$.

(iii) *If $I(F_1) < +\infty$, the equality $I(F_1) = I(F)$ holds if and only if $F = (F_1)^{\otimes j}$.*

PROOF OF LEMMA 3.7.

Proof of (i). If $I(F) = +\infty$ the conclusion is clear. Otherwise, thanks to the equivalent definition $I^{(3)}$ of the Fisher information and the symmetry assumption of F , we have

$$\begin{aligned} I(F) &= \sup_{\psi \in C_b(E^j)^{dj}} \frac{1}{j} \int_{E^j} \left(\psi(x_1, \dots, x_j) \cdot \nabla F - F \frac{|\psi(x_1, \dots, x_j)|^2}{4} \right) \\ &= \sup_{\psi \in C_b(E^j)^d} \int_{E^j} \left(\psi(x_1, \dots, x_j) \cdot \nabla_1 F - F \frac{|\psi(x_1, \dots, x_j)|^2}{4} \right) \\ &\geq \sup_{\psi \in C_b(E^\ell)^d} \int_{E^j} \left(\psi(x_1, \dots, x_\ell) \cdot \nabla_1 F - F \frac{|\psi(x_1, \dots, x_\ell)|^2}{4} \right) \\ &= \sup_{\psi \in C_b(E^\ell)^d} \int_{E^\ell} \left(\psi \cdot \nabla_1 F_\ell - F_\ell \frac{|\psi|^2}{4} \right) = I(F_\ell). \end{aligned}$$

Proof of the superadditivity property (ii). The first proof of that result seems to be the one by Carlen in [16, Theorem 3]. We sketch now another proof that uses the third formulation $I^{(3)}$. We recall that in the definition of $I_j^{(3)}(F)$ the supremum is taken over the $\psi = (\psi_1, \dots, \psi_j)$, with all $\psi_i : E^j \rightarrow \mathbb{R}^d$. We now restrict the supremum over the ψ such that:

- The ℓ first ψ_i depend only on (x_1, \dots, x_ℓ) , with the notation $\psi^\ell = (\psi_1, \dots, \psi_\ell)$.
- The $(j-\ell)$ last ψ_i depend only on $(x_{\ell+1}, \dots, x_j)$, with the notation $\psi^{j-\ell} = (\psi_{\ell+1}, \dots, \psi_j)$.

We then have the inequality

$$\begin{aligned}
I_j(F) &\geq \sup_{\psi^\ell, \psi^{j-\ell}} \int_{E^j} [\nabla_\ell f \cdot \psi^\ell + \nabla_{j-\ell} f \cdot \psi^{j-\ell} - f \frac{|\psi^\ell|^2 + |\psi^{j-\ell}|^2}{4}] \\
&= \sup_{\psi^\ell \in C_b^1(E^\ell)^{\ell d}} \int_{E^\ell} [\nabla f_\ell \cdot \psi^\ell - f_\ell \frac{|\psi^\ell|^2}{4}] \\
&\quad + \sup_{\psi^{j-\ell} \in C_b^1(E^{j-\ell})^{(j-\ell)d}} \int_{E^{j-\ell}} [\nabla f_{j-\ell} \cdot \psi^{j-\ell} - f_{j-\ell} \frac{|\psi^{j-\ell}|^2}{4}] \\
&= I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell})
\end{aligned}$$

If the inequality is an equality, we use the remark made at the end of Step 1 in the proof of Lemma 3.5 : Maximizing sequences ψ_n^ℓ and $\psi_n^{j-\ell}$ for respectively I_ℓ (resp. $I_{j-\ell}$) should converge pointwise towards $2 \nabla \ln f_\ell$ (resp. $2 \nabla \ln f_{j-\ell}$) up to some subsequence, a.e. on $\{f_\ell \neq 0\}$ (resp. $\{f_{j-\ell} \neq 0\}$). If we have equality, we also must have $(\psi_n^\ell, \psi_n^{j-\ell}) \rightarrow 2 \nabla \ln f$ on $\{f \neq 0\}$, a set that is included in $\{f_\ell \neq 0\} \times \{f_{j-\ell} \neq 0\}$ and thus

$$\nabla \ln f = (\nabla \ln f_\ell, \nabla \ln f_{j-\ell}) = \nabla \ln(f_\ell \otimes f_{j-\ell}),$$

which implies the claimed equality since f and $f_\ell \otimes f_{j-\ell}$ are probability measures.

The case of equality (iii). Using recursively the superadditivity in that particular case, we get with the notation $F_1 = f$

$$I(f) = I(F) \geq \frac{j-1}{j} I(F_{j-1}) + \frac{1}{j} I(f) \geq \frac{j-2}{j} I(F_{j-2}) + \frac{2}{j} I(f) \geq \dots \geq I(f).$$

Therefore, all the inequalities are equalities. We obtain that

$$F = F_{j-1} \otimes f = F_{j-2} \otimes f \otimes f = \dots = f^{\otimes j},$$

by applying recursively the case of equality in (3.14). \square

It is classical and essentially a consequence of the Sobolev inequality and the Rellich-Kondrachov Theorem (together with very standard manipulations on the entropy functional which are similar to the ones presented at the end of the proof of Theorem 4.13) that for (f_n) a sequence of $\mathbf{P}(E)$, the conditions

$$f_n \rightharpoonup f \text{ weakly in } \mathbf{P}(E), \quad M_k(f_n) \text{ bounded, } k > 0, \quad \text{and} \quad I(f_n) \leq C$$

imply that $H(f_n) \rightarrow H(f)$. A natural question is whether a similar result holds for a sequence (F^N) in $\mathbf{P}(E^N)$. Before answering affirmatively to that question, we establish a normalized non-relative HWI inequality for a large class of sets $E \subset \mathbb{R}^d$. It is a variant of the famous HWI inequality of Otto-Villani [61] that will be the cornerstone of the argument. Let us mention that its good behaviour in any dimension is of particular importance here and it is due to the good (separate) behaviours of H , W_2 and I with respect to the dimension.

Proposition 3.8. *Assume that $E \subset \mathbb{R}^d$ is a bi-Lipschitz volume preserving deformation of a convex set of \mathbb{R}^d , $d \geq 1$: there exists a convex subset $E_1 \subset \mathbb{R}^d$ and a bi-lipschitz diffeomorphism $T : E_1 \rightarrow E$ which preserves the volume (i.e. its Jacobian is always equal to 1). Then, the normalized non relative HWI inequality holds in E : there exists a constant $C_E \in [1, \infty)$ such that*

$$(3.15) \quad \forall F^N, G^N \in \mathbf{P}_2(E^N) \quad H(F^N) \leq H(G^N) + C_E W_2(F^N, G^N) \sqrt{I(F^N)}.$$

More precisely, the above inequality holds with $C_E := \|\nabla T\|_\infty \|\nabla T^{-1}\|_\infty$ where $\|\nabla T\|_\infty := \sup_{v \in E} \sup_{|h|_2 \leq 1} |\nabla T(v) h|_2$.

Before going to the proof, remark that the class of set E which are bi-Lipschitz volume preserving deformation of convex set is rather large. For instance, it is shown in [31, Theorem 5.4] that any star-shaped bounded domain with Lipschitz boundary (and some additional assumptions) is in the previously mentioned class.

PROOF OF PROPOSITION 3.8. We proceed in three steps.

Step 1. $E = \mathbb{R}^d$. Let us first recall the famous HWI inequality of Otto-Villani. Consider $\rho = e^{-V(x)} dx$ a probability measure on \mathbb{R}^D such that $D^2V \geq 0$. For any probability measures $f_0, f_1 \in \mathbf{P}_2(\mathbb{R}^D)$, there holds

$$(3.16) \quad H_D(f_0|\rho) \leq H_D(f_1|\rho) + \tilde{W}_2(f_0, f_1) \sqrt{I_D(f_0|\rho)},$$

where H_D and I_D stand for the non normalized relative entropy and relative Fisher information defined in (3.4) and (3.12) respectively, and \tilde{W}_2 stands for the non normalized quadratic MKW distance in \mathbb{R}^D based on the usual Euclidean norm $|V| = (\sum_{i=1}^D |v_i|^2)^{1/2}$ for any $V = (v_1, \dots, v_D) \in \mathbb{R}^D$. Inequality (3.16) has been proved in [61], see also [72, 73, 60, 9, 23]. We easily deduce the “non relative” inequality (3.15) from the “relative” inequality (3.16). In order to do so, we simply apply the HWI inequality (3.16) in \mathbb{R}^D , $D = dN$, with respect to the Gaussian $\gamma_\lambda(v) := (2\pi\lambda)^{-D/2} e^{-|v|^2/2\lambda}$, and we get

$$H_D(F^N|\gamma_\lambda) \leq H_D(G^N|\gamma_\lambda) + \tilde{W}_2(F^N, G^N) \sqrt{I_D(F^N|\gamma_\lambda)}.$$

We write the relative entropy and the relative Fisher information in terms of the non-relative ones, and we get

$$\begin{aligned} H_D(F^N|\gamma_\lambda) &= H_D(F^N) - \int F^N \ln(\gamma_\lambda) = H_D(F^N) + \frac{D}{2} \log(2\pi\lambda) + \frac{M_2(F^N)}{2\lambda}, \\ I_D(F^N|\gamma_\lambda) &= \int F^N \left| \nabla \ln F^N + \frac{v}{\lambda} \right|^2 = I_D(f_0) + \frac{2}{\lambda} \int v \cdot \nabla f_0 + \frac{M_2(f_0)}{\lambda^2} \\ &= I_D(f_0) - \frac{2D}{\lambda} + \frac{M_2(f_0)}{\lambda^2}. \end{aligned}$$

Inserting this in the relative HWI inequality, simplifying the terms involving $\log(2\pi\lambda)$, letting $\lambda \rightarrow +\infty$ and dividing the resulting limit by N , we obtain the claimed result.

Step 2. $E \subset \mathbb{R}^d$ is convex. The proof is the same as in the case $E = \mathbb{R}^d$ using that the HWI inequality (3.16) holds in a convex set. We have no precise reference for that last result but all the necessary arguments can be find in [73]. More precisely, [73, Chapter 20] explains that the HWI inequality (3.16) holds when the entropy is displacement convex, while it is proved in [73, Chapters 16 and 17] that the entropy on a convex set E is displacement convex, exactly as on \mathbb{R}^d .

Step 3. General case. We choose two absolutely continuous probability measures F^N and G^N on E^N , and defined the corresponding probability measures F_1^N and G_1^N on E_1^N by

$$F_1^N(v_1, \dots, v_N) := F^N(T(v_1), \dots, T(v_N)) = F^N \circ T^{\otimes N}(V),$$

and the same formula for G_1^N . It can be checked that $\nabla_{v_j} F_1^N = {}^t \nabla T(v_j) \nabla_{v_j} F^N \circ T^{\otimes N}$, so that $|\nabla_{v_j} F_1^N| \leq \|\nabla T\|_\infty |\nabla_{v_j} F^N \circ T^{\otimes N}|$. Turning to Fisher information, it comes

$$I(F_1^N) := \int_{E_1^N} \frac{|\nabla F_1^N|^2}{F_1^N} dV \leq \|\nabla T\|_\infty^2 \int_{E_1^N} \frac{|\nabla F^N \circ T^{\otimes N}|^2}{F^N \circ T^{\otimes N}} dV = \|\nabla T\|_\infty^2 I(F^N),$$

where we have used the fact that T preserves the volume.

For the MKW distance, remark that $|(T^{-1})^{\otimes N}(V) - (T^{-1})^{\otimes N}(V')| \leq \|\nabla T^{-1}\|_\infty |V - V'|$. Therefore,

$$\begin{aligned}
W_2(F_1^N, G_1^N)^2 &= \inf_{\pi_1 \in \Pi(F_1^N, G_1^N)} \int |V - V'|^2 \pi_1(dV, dV') \\
&= \inf_{\pi \in \Pi(F^N, G^N)} \int |(T^{-1})^{\otimes N}(V) - (T^{-1})^{\otimes N}(V')|^2 \pi(dV, dV') \\
&\leq \|\nabla T^{-1}\|_\infty^2 \inf_{\pi \in \Pi(F^N, G^N)} \int |V - V'| \pi(dV, dV') \\
&= \|\nabla T^{-1}\|_\infty^2 W_2(F^N, G^N)^2.
\end{aligned}$$

For the entropy, the preservation of volume ensures the equality $H(F_1^N) = H(F^N)$, and a similar one for G^N . Finally, using the HWI inequality in E_1 proved in step 2 and the above properties, we get

$$\begin{aligned}
H(F^N) &= H(F_1^N) \leq H(G_1^N) + \sqrt{I(F_1^N)} W_2(F_1^N, G_1^N) \\
&\leq H(G^N) + \|\nabla T\|_\infty \|\nabla T^{-1}\|_\infty \sqrt{I(F^N)} W_2(F^N, G^N),
\end{aligned}$$

which is exactly the claimed result. \square

Let us finally prove now our main result Theorem 1.4 which is a consequence of the characterization of the Kac's chaos in Theorem 2.4 together with Proposition 3.8.

PROOF OF THEOREM 1.4. We recall that the implication (iii) \Rightarrow (iv) has been yet proven in Theorem 3.4. We split the proof into two steps.

Step 1. (i) \Rightarrow (ii). Fix a $j \in \mathbb{N}$, there exists a subsequence of (G^N) , still denoted by (G^N) , and some compatible and symmetric probability measures $F_j \in \mathbf{P}(E^j)$, such that $G_j^N \rightarrow F_j$ weakly in $\mathbf{P}(E^j)$. In particular $F_1 = f$. As a consequence of Lemma 3.5 and Lemma 3.7 point (i), we have

$$I(f) \leq I(F_j) \leq \liminf I(G_j^N) \leq \liminf I(G^N) = I(f).$$

Using now the third point of Lemma 3.7 we deduce $F_j = f^{\otimes j}$. The uniqueness of the limit implies that the whole sequence G^N is in fact f -Kac's chaotic.

Step 2. (ii) \Rightarrow (iii). We write twice the normalized non relative HWI inequality of Proposition 3.8, and get

$$|H(G^N) - H(f^{\otimes N})| \leq C_E W_2(G^N, f^{\otimes N}) \left(\sqrt{I(G^N)} + \sqrt{I(f^{\otimes N})} \right).$$

Using the previous inequalities together with the inequality of the Lemma 2.2

$$W_2(G^N, f^{\otimes N}) \leq C_E 2^{\frac{3}{2}} [M_k(G_1^N) + M_k(f)]^{1/k} W_1(G^N, f^{\otimes N})^{1/2-1/k}$$

we get (1.8) since $M_k(f) \leq \sup M_k(G_1^N)$ and $I(f) \leq \sup I(G^N)$. \square

4. PROBABILITY MEASURES ON THE “KAC’S SPHERES”

We generalize the preceding two sections to the important case of probability measures with support on the “Kac’s spheres”

$$\mathcal{KS}_N := \{V = (v_1, \dots, v_N) \in \mathbb{R}^N, v_1^2 + \dots + v_N^2 = N\}.$$

We refer to [19] where similar results are obtained to the (even more important) case of probability measures with support on the “Boltzmann’s spheres”

$$\mathcal{BS}_N := \{V = (v_1, \dots, v_N) \in (\mathbb{R}^3)^N, |v_1|^2 + \dots + |v_N|^2 = N, v_1 + \dots + v_N = 0\}.$$

4.1. On uniform probability measures on the Kac’s spheres as $N \rightarrow \infty$.

Definition 4.1. For any $N \in \mathbb{N}^*$ and $r > 0$, we denote by $\sigma^{N,r}$ the uniform probability measure of \mathbb{R}^N carried by the sphere S_r^{N-1} defined by

$$S_r^{N-1} := \{V \in \mathbb{R}^N; |V|^2 = r^2\}.$$

We define $\sigma^N \in \mathbf{P}(E^N)$, $E = \mathbb{R}$, the sequence $\sigma^N := \sigma^{N, \sqrt{N}}$ of probability measures uniform on the Kac’s spheres

$$\mathcal{KS}_N := S_{\sqrt{N}}^{N-1} := \{V \in \mathbb{R}^N; |V|^2 = N\}.$$

We begin with a classical and elementary lemma that we will use several times in the sequel.

Lemma 4.2. (i) For any $1 \leq \ell \leq N-1$, there holds

$$\sigma_\ell^N(V) = \left(1 - \frac{|V|^2}{N}\right)_+^{\frac{N-\ell-2}{2}} \frac{|S_1^{N-\ell-1}|}{N^{\ell/2} |S_1^{N-1}|},$$

where we recall that $|S_1^{k-1}| = 2\pi^{k/2}/\Gamma(k/2)$.

(ii) For any fixed ℓ , the sequence $(\sigma_\ell^N)_{N \geq N_\ell}$ is bounded in L^∞ (with $N_\ell = \ell + 4$), in H^s for any $s \geq 0$ (with $N_\ell = N(\ell, k)$ large enough) and the exponential moment $M_{2,1/4}(\sigma_1^N)$ defined in (2.27) is bounded (uniformly in N).

(iii) For any function $\varphi \in C_b(\mathbb{R}^N)$, any $r > 0$ and $1 \leq \ell \leq N-1$, there holds

$$\int_{S_r^{N-1}} \varphi(V, V') d\sigma_r^N(V, V') = \int_{B^\ell(r)} \frac{|S_1^{N-\ell-1}|}{|S_r^{N-1}|} \left\{ \int_{S_{\sqrt{r^2-V^2}}^{N-\ell-1}} \varphi(V, V') d\sigma_{\sqrt{r^2-V^2}}^{N-\ell}(V') \right\} dV,$$

where $V \in \mathbb{R}^\ell$ and $V' \in \mathbb{R}^{N-\ell}$. This precisely means that

$$\sigma^N(dV, dV') = \sigma_\ell^N(dV) \sigma_{\sqrt{N-|V|^2}}^{N-\ell}(dV').$$

PROOF OF LEMMA 4.2. (i) One possible definition of $\sigma^{N,r}$ is

$$\sigma^{N,r} := \frac{1}{r^{N-1} |S_1^{N-1}|} \lim_{h \rightarrow 0} \frac{1}{h} \left(\mathbf{1}_{B^N(r+h)} - \mathbf{1}_{B^N(r)} \right), \quad B^N(\rho) := \{V \in \mathbb{R}^N; |V| \leq \rho\},$$

where the surface $r^{N-1} |S_1^{N-1}|$ of the Sphere S_r^{N-1} stands for the normalization constant such that $\sigma^{N,r}$ is a probability measure. For any $\varphi \in C_b(E^\ell)$, $1 \leq \ell \leq N-1$, we compute

$$\begin{aligned} \langle \mathbf{1}_{B(\rho)}, \varphi \otimes \mathbf{1}^{N-\ell} \rangle &= \int_{\mathbb{R}^\ell} \mathbf{1}_{|V|^2 \leq \rho^2} \varphi(V) \left\{ \int_{\mathbb{R}^{N-\ell}} \mathbf{1}_{x_{\ell+1}^2 + \dots + x_N^2 \leq \rho^2 - |V|^2} dx_{\ell+1} \dots dx_N \right\} dV \\ &= \int_{\mathbb{R}^\ell} \varphi(V) \omega^{N-\ell} (\rho^2 - |V|^2)_+^{\frac{N-\ell}{2}} dV, \end{aligned}$$

where $\omega^k = |B^k(1)|$ is the volume of the unit ball of \mathbb{R}^k . We deduce

$$\sigma_\ell^N(r) = \frac{1}{Z_{N,r}} \frac{d}{dr} \left[\omega^{N-\ell} (r^2 - |V|^2)_+^{\frac{N-\ell}{2}} \right] = \frac{\omega^{N-\ell} (N-\ell)}{r^{N-1} |S^{N-1}|} r (r^2 - |V|^2)_+^{\frac{N-\ell-2}{2}}.$$

We conclude using the relation $|S_1^{k-1}| = k \omega^k$.

(ii) The estimates on σ_ℓ^N are deduced from its explicit expression after some tedious but easy calculations. We only prove the last one which will be a key argument in the proof of the accurate rate of chaoticity in Theorem 1.5. For any $k \geq 1$ and introducing $n := (N-4)/2$, we easily estimate

$$\begin{aligned} \int_{\mathbb{R}^2} |v_1|^{2k} \sigma_2^N(dv) &= \frac{1}{2\pi} \frac{N-2}{N} \int_{\mathbb{R}^2} |v_1|^K \left(1 - \frac{|v|^2}{N}\right)_+^{\frac{N-4}{2}} dv \\ &\leq \int_0^{\sqrt{N}} r^{K+1} \left(1 - \frac{r^2}{N}\right)^{\frac{N-4}{2}} dr \\ &= N^{k+1} \int_0^1 s^k (1-s)^n ds. \end{aligned}$$

Thanks to $k+1$ integrations by parts, we deduce

$$\begin{aligned} \int_{\mathbb{R}^2} |v_1|^{2k} \sigma_2^N(dv) &\leq N^{k+1} \int_0^1 (1-z)^k z^n dz \\ &= N^{k+1} \frac{k}{n+1} \int_0^1 (1-z)^{k-1} z^{n+1} dz \\ &= N^{k+1} \frac{k}{n+1} \cdots \frac{2}{n+k-1} \frac{1}{n+k} \frac{1}{n+k+1}, \end{aligned}$$

and then

$$\begin{aligned} \int_{\mathbb{R}^2} e^{|v|^2/4} \sigma_1^N(v) dv &\leq \sum_{k=0}^{\infty} \frac{1}{k! 4^k} \int_{\mathbb{R}^2} |v_1|^{2k} \sigma_2^N(dv) \\ &= \sum_{k=0}^{\infty} \frac{1}{4^k} \frac{(2n+4)^{k+1}}{(n+1) \cdots (n+k+1)} \\ &\leq 2 \sum_{k=0}^{\infty} \frac{1}{2^k} \frac{(n+2)}{(n+1)} \leq 6. \end{aligned}$$

(iii) We come back to the proof of (i). We set $m = \ell$ and $n = N - \ell$ and we write

$$\begin{aligned}
\langle \sigma^{N,r}, \varphi \rangle &= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{B^N(r+h)} \varphi - \int_{B^N(r)} \varphi \right] \\
&= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{|v| \leq r+h} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi - \int_{|v| \leq r} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi \right] \\
&\quad + \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{|v| \leq r} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi - \int_{|v| \leq r} \int_{|v'| \leq \sqrt{r^2 - |v|^2}} \varphi \right] \\
&= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \int_{r \leq |v| \leq r+h} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi \\
&\quad + \frac{1}{Z_{N,r}} \int_{B_r^m} \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{B^n(\sqrt{(r+h)^2 - |v|^2})} \varphi - \int_{B^n(\sqrt{r^2 - |v|^2})} \varphi \right].
\end{aligned}$$

We invert the integral and the limit on the last line using dominated convergence, since the integral on v' are bounded by $\|\varphi\|_\infty / \sqrt{r^2 - |v|^2}$. The first term is bounded (for any $0 < h \leq r$) by

$$\frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \int_{r \leq |v| \leq r+h} \int_{|v'| \leq \sqrt{3rh}} |\varphi| \leq C_{N,r} \|\varphi\|_{L^\infty} \lim_{h \rightarrow 0} \sqrt{h} = 0,$$

and the second term converges to

$$\int_{B^m(r)} \frac{Z_{n, \sqrt{r^2 - |v|^2}}}{Z_{m+n,r}} \left\{ \int_{S_{\sqrt{r^2 - |v|^2}}^{n-1}} \varphi(v, v') d\sigma_{\sqrt{r^2 - |v|^2}}^n(v') \right\} dv,$$

which is exactly the claimed identity. \square

Let us recall the following classical result.

Theorem 4.3. *The sequence σ^N is γ -chaotic, where γ still stands for the gaussian distribution $\gamma(dx) = (2\pi)^{-1/2} e^{-x^2/2} dx$ on \mathbb{R} , and more precisely*

$$(4.1) \quad \|\sigma_\ell^N - \gamma^{\otimes \ell}\|_{L^1} \leq 2 \frac{\ell + 3}{N - \ell - 3} \quad \text{pour tout } 1 \leq \ell \leq N - 4.$$

The fact that σ^N is γ -chaotic is sometime called ‘‘Poincaré’s Lemma’’. In fact, it should go back to Mehler [51] in 1866. Anyway, we refer to [25, 17] for a bibliographic discussion about this important result, and to [25] for a proof of estimate (4.1). We give now a different quantitative version of the ‘‘Poincaré’s Lemma’’.

Theorem 4.4. *There exists a numerical constant $C \in (0, \infty)$ such that*

$$(4.2) \quad \Omega_N(\sigma^N; \gamma) := W_1(\sigma^N, \gamma^{\otimes N}) \leq \frac{C}{\sqrt{N}}.$$

Remark 4.5. *It is worth observing that it is not clear that one can deduce (4.2) from (4.1) or that the reverse implication holds. In particular, using (4.1) and Theorem 2.4 we obtain an estimate on $W_1(\sigma^N, \gamma^{\otimes N})$ which is weaker than (4.2).*

PROOF OF THEOREM 4.4. There is a simple transport map from $\gamma^{\otimes N}$ onto σ^N which is given by the radial projection $P : V \mapsto \frac{V}{|V|_2}$ with the notation $|V|_k = (N^{-1} \sum_i |v_i|^k)^{1/k}$ for any $k > 0$ for the normalized distance of order k . The fact it is an admissible map comes from the invariance by rotation of $\gamma^{\otimes N}$ and σ^N . Is it optimal? It is not obvious

because $P(V)$ is not necessary the point of \mathcal{KS}_N which is the closest to $V \in \mathbb{R}^N$, for the $|\cdot|_1$ distance (for which it costs less to displace in the direction of the axis). However, it may still be optimal for rotational symmetry reasons, but it is less obvious. Nevertheless, it will be sufficient for our estimate. Since,

$$|P(V) - V|_1 = \left| \frac{1}{|V|_2} - 1 \right| |V|_1$$

we get as all our distances are normalized

$$\begin{aligned} W_1(\gamma^{\otimes N}, \sigma^N) &\leq \int_{\mathbb{R}^N} |P(V) - V|_1 \gamma^{\otimes N}(dV) \\ &= \int_{\mathbb{R}^N} \left| \frac{1}{|V|_2} - 1 \right| |V|_1 \gamma^{\otimes N}(dV) \\ &= \left(\int_0^{+\infty} \left| \frac{\sqrt{N}}{R} - 1 \right| R^N e^{-R^2/2} dR \right) \frac{|S^{N-1}|}{(2\pi)^{N/2}} \left(\int_{S_1^{N-1}} |V|_1 d\sigma^{N,1} \right). \end{aligned}$$

Using that $|V|_1 \leq |V|_2$ because of the normalization, we may bound the last integral by

$$\int_{S_1^{N-1}} |V|_1 d\sigma^N \leq \int_{S_1^{N-1}} |V|_2 d\sigma^N = N^{-1/2}.$$

Remark that this integral is also equal to $\frac{1}{\sqrt{N}} M_1(\sigma^N)$ which can be explicitated thanks to the formula for σ_1^N of Lemma 4.2. Using this in the previous inequality and performing the change of variable $R = \sqrt{N}R'$, we get

$$\begin{aligned} W_1(\gamma^{\otimes N}, \sigma^N) &\leq \frac{|S^{N-1}|}{\sqrt{N}(2\pi)^{N/2}} \int_0^{+\infty} |\sqrt{N} - R| R^{N-1} e^{-R^2/2} dR \\ &\leq \frac{|S^{N-1}| N^{\frac{N}{2}}}{(2\pi)^{N/2}} \int_0^{+\infty} |1 - R'| (R')^{N-1} e^{-NR'^2/2} dR'. \end{aligned}$$

We can simplify the prefactor, using the formula for $|S^{N-1}|$ and Stirling's formula

$$\begin{aligned} \frac{|S^{N-1}| N^{\frac{N}{2}}}{(2\pi)^{N/2}} &= \frac{N^{\frac{N}{2}}}{\Gamma(\frac{N}{2}) 2^{N/2-1}} \\ &= \frac{\sqrt{N} e^{N/2}}{\sqrt{\pi}} [1 + O(1/N)]. \end{aligned}$$

Turning back to the transportation cost, we get

$$W_1(\gamma^{\otimes N}, \sigma^N) \leq \frac{e\sqrt{N}}{\sqrt{\pi}} [1 + O(1/N)] \int_0^{+\infty} \left(R e^{(1-R^2)/2} \right)^{N-1} e^{-R^2/2} |1 - R| dR.$$

After studying the function $g(r) = r e^{(1-r^2)/2}$, we remark that it is strictly increasing from 0 to 1, then strictly decreasing from 1 to $+\infty$, that its maximum in 1 is 1, and that $g(1 + \varepsilon) = 1 - \varepsilon^2 + O(\varepsilon^3)$. We shall also use the less sharp but exact bound

$$g(1 + \varepsilon) \leq 1 - \frac{\varepsilon^2}{4}, \quad \text{for } \varepsilon \in [-\frac{1}{2}, \sqrt{2} - 1].$$

We can now cut the previous integral in three parts $\int_0^{1/2} + \int_{1/2}^{\sqrt{2}} + \int_{\sqrt{2}}^{+\infty}$. We bound the first part by

$$\int_0^{1/2} \dots \leq \frac{1}{2} g\left(\frac{1}{2}\right)^{N-1},$$

and the third part by

$$\int_{\sqrt{2}}^{+\infty} \dots \leq g(\sqrt{2})^{N-1} \int_{\sqrt{2}}^{+\infty} e^{-r^2/2} r dr = g(\sqrt{2})^{N-1} e^{-1}.$$

For the last part, we perform the change of variable $r = 1 + u/\sqrt{N}$. It comes

$$\begin{aligned} \int_{1/2}^{\sqrt{2}} \dots &= \frac{1}{N} \int_{-\sqrt{N}/2}^{(\sqrt{2}-1)\sqrt{N}} g\left(1 + \frac{u}{\sqrt{N}}\right)^{N-1} |u| e^{-u/\sqrt{N} - u^2/2N} du \\ &\leq \frac{1}{N} \int_{-\sqrt{N}/2}^{(\sqrt{2}-1)\sqrt{N}} \left(1 - \frac{u^2}{4N}\right)^{N-1} |u| du \\ &\leq \frac{1}{N} (1 + O(N^{-1})) \int_{-\infty}^{+\infty} e^{-\frac{u^2}{4}} |u| du \\ &\leq \frac{4}{N} (1 + O(N^{-1})) \end{aligned}$$

Putting all together, we finally get

$$W_1(\gamma^{\otimes N}, \sigma^N) \leq \frac{C}{\sqrt{N}} (1 + O(N^{-1})) + C\sqrt{N}\lambda^N,$$

with $\lambda = \max(g(\sqrt{2}), g(1/2)) < 0.86$. This implies the claimed inequality. \square

PROOF OF (1.9) IN THEOREM 1.5. The proof of the last estimate in (1.9) follows from (4.2) and Lemma 4.2-(ii) together with (2.29). \square

4.2. Conditioned tensor products on the Kac's spheres. We begin with a sharp version of the local central limit theorem (local CLT) or Berry-Esseen type theorem which will be the cornerstone argument in this section.

Theorem 4.6. *Consider $g \in \mathbf{P}_3(\mathbb{R}^D) \cap L^p(\mathbb{R}^D)$, $p \in (1, \infty]$, such that*

$$(4.3) \quad \int_{\mathbb{R}^D} x g(x) dx = 0, \quad \int_{\mathbb{R}^D} x \otimes x g(x) dx = Id, \quad \int_{\mathbb{R}^D} |x|^3 g(x) dx =: M_3.$$

We define the iterated and renormalized convolution by

$$(4.4) \quad g_N(x) := \sqrt{N} g^{(*N)}(\sqrt{N} x).$$

There exists an integer $N(p)$ and a constant $C_{BE} = C(p, k, M_3(g), \|g\|_{L^p})$ such that

$$(4.5) \quad \forall N \geq N(p) \quad \|g_N - \gamma\|_{L^\infty} \leq \frac{C_{BE}}{\sqrt{N}}.$$

Remark 4.7. *Theorem 4.6 is a sharper but less general version of [17, Proposition 26]. The proof follows the proof of [17, Proposition 26] and uses an argument from [17, Proposition 26], see also [45]. The first local CLT have been established in the pioneer works by A. C. Berry [7] and C.-G. Esseen [29] who proved the convergence in $\mathcal{O}(1/\sqrt{N})$ uniformly on the distribution function in dimension $D = 1$, see for instance [30, Theorem 5.1, Chapter XVI]. Since that time, many variants of the local CLT have been established*

corresponding to different regularity assumption made on the probability measure g , we refer the interested reader to the recent works [63], [8], [3] and the references therein.

The proof of Theorem 4.6 use the following technical lemma which proof is postponed after the proof of the Theorem.

Lemma 4.8. (i) Consider $g \in \mathbf{P}_3(\mathbb{R}^D)$ satisfying (4.3). There exists $\delta \in (0, 1)$ such that

$$\forall \xi \in B(0, \delta) \quad |\hat{g}(\xi)| \leq e^{-|\xi|^2/4}.$$

(ii) Consider $g \in \mathbf{P}(\mathbb{R}^D) \cap L^p(\mathbb{R}^D)$, $p \in (1, \infty]$. For any $\delta > 0$ there exists $\kappa = \kappa(M_3(g), \|g\|_{L^p}, \delta) \in (0, 1)$ such that

$$(4.6) \quad \sup_{|\xi| \geq \delta} |\hat{g}(\xi)| \leq \kappa(\delta).$$

PROOF OF THEOREM 4.6. We follow closely the proof of [17, Theorem 27] which is more general but less precise, and we use a trick that we found in the proof of [34, Theorem 1]. We observe that

$$\hat{g}_N(\xi) = (\hat{g}(\xi/\sqrt{N}))^N, \quad \hat{\gamma}(\xi) = (\hat{\gamma}(\xi/\sqrt{N}))^N.$$

Because $g \in L^1 \cap L^p$, the Hausdorff-Young inequality implies $\hat{g} \in L^{p'} \cap L^\infty$ with $p' \in [1, \infty)$, and then $\hat{g}_N(\xi) = (\hat{g}(\xi/\sqrt{N}))^N \in L^1$ for any $N \geq p'$. As a consequence we may write

$$|g_N(x) - \gamma(x)| = (2\pi)^D \left| \int_{\mathbb{R}^D} (\hat{g}_N(\xi) - \hat{\gamma}(\xi)) e^{i\xi \cdot x} d\xi \right| \leq (2\pi)^D \int_{\mathbb{R}^D} |\hat{g}_N - \hat{\gamma}| d\xi.$$

We split the above integral between low and high frequencies

$$\begin{aligned} \|g_N - \gamma\|_{L^\infty} &\leq \int_{|\xi| \geq \sqrt{N}\delta} |\hat{g}_N| d\xi + \int_{|\xi| \geq \sqrt{N}\delta} |\hat{\gamma}| d\xi \\ &\quad + \int_{|\xi| < \sqrt{N}\delta} |\hat{g}_N - \hat{\gamma}| d\xi \quad (=: T_1 + T_2 + T_3). \end{aligned}$$

For the first term, we have

$$\begin{aligned} T_1 &\leq \int_{|\xi| \geq \sqrt{N}\delta} \left| \hat{g} \left(\frac{\xi}{\sqrt{N}} \right) \right|^N d\xi = N^{d/2} \int_{|\eta| \geq \delta} |\hat{g}(\eta)|^N d\eta \\ &\leq \left(\sup_{|\eta| \geq \delta} |\hat{g}(\eta)| \right)^{N-p'} N^{d/2} \int_{|\eta| \geq \delta} |\hat{g}(\eta)|^{p'} d\eta \\ &\leq \kappa(\delta)^{N-p'} N^{d/2} C_p \|g\|_{L^p}^p \end{aligned}$$

with $\delta \in (0, 1)$ given by point (i) of Lemma 4.8, $\kappa(\delta)$ given by point (ii) of Lemma 4.8 and $N \geq p'$. The second term may be estimated in the same way, and we clearly obtain that there exists a constant $C_1 = C_1(D, p, \|g\|_{L^p})$ such that

$$(4.7) \quad T_1 + T_2 \leq \frac{C_1}{\sqrt{N}}.$$

Concerning the third term, we write

$$T_3 = \int_{|\xi| \leq \sqrt{N}\delta} \frac{|\hat{g}_N(\xi) - \hat{\gamma}_N(\xi)|}{|\xi|^3} |\xi|^3 d\xi,$$

with

$$\begin{aligned}
\frac{|\hat{g}_N(\xi) - \hat{\gamma}_N(\xi)|}{|\xi|^3} &= \frac{1}{N^{3/2}} \frac{|\hat{g}(\xi/\sqrt{N})^N - \hat{\gamma}(\xi/\sqrt{N})^N|}{|\xi/\sqrt{N}|^3} \\
&= \frac{1}{N^{3/2}} \frac{|\hat{g}(\xi/\sqrt{N}) - \hat{\gamma}(\xi/\sqrt{N})|}{|\xi/\sqrt{N}|^3} \\
&\quad \times \left| \sum_{k=0}^{N-1} \hat{g}(\xi/\sqrt{N})^k \hat{\gamma}(\xi/\sqrt{N})^{N-k-1} \right|.
\end{aligned}$$

Estimate (i) of Lemma 4.8 implies

$$\begin{aligned}
&\left| \sum_{k=0}^{N-1} \hat{g}(\xi/\sqrt{N})^k \hat{\gamma}(\xi/\sqrt{N})^{N-k-1} \right| \\
&\leq \sum_{k=0}^{N-1} e^{-\frac{|\xi|^2}{4N} k} e^{-\frac{|\xi|^2}{2N} (N-k-1)} \leq N e^{-\frac{|\xi|^2}{4} \frac{N-1}{N}} \leq N e^{-\frac{|\xi|^2}{8}}.
\end{aligned}$$

We deduce

$$\begin{aligned}
T_3 &= \frac{1}{N^{3/2}} \left(\sup_{\eta} \frac{|\hat{g}(\eta) - \hat{\gamma}(\eta)|}{|\eta|^3} \right) \int_{\mathbb{R}^D} N e^{-\frac{|\xi|^2}{8}} |\xi|^3 d\xi \\
&\leq \frac{1}{N^{1/2}} (M_3(g) + M_3(\gamma)) C_{k,d}.
\end{aligned}$$

We conclude by gathering the estimates on each term. \square

PROOF OF LEMMA 4.8. Thanks to a Taylor expansion, we have

$$\begin{aligned}
\hat{g}(\xi) &= 1 - \frac{\xi^2}{2} + \mathcal{O}(M_3(g) |\xi|^3) \\
\hat{\omega}(\xi) &= 1 - \frac{\xi^2}{4} + \mathcal{O}(|\xi|^3), \quad \omega(x) := \frac{1}{\sqrt{\pi}} e^{-x^2},
\end{aligned}$$

from which we deduce that there exists $\delta = \delta(M_3(g)) \in (0, 1)$ small enough such that

$$\forall \xi \in B_\delta \quad |\hat{g}(\xi)| \leq 1 - \frac{3}{8} \xi^2 \leq \hat{\omega}(\xi), \quad \hat{\omega}(\xi) := e^{-\xi^2/4}.$$

That is nothing but (i). On the other hand, (ii) is a consequence of [17, Proposition 26, (iii)]. \square

For a given “smooth enough” probability measure $f \in \mathbf{P}(E)$, $E = \mathbb{R}$, we define

$$Z_N(r) := \int_{S^{N-1}(r)} f^{\otimes N} d\sigma^{N,r}, \quad Z'_N(r) := \int_{S^{N-1}(r)} \frac{f^{\otimes N}}{\gamma^{\otimes N}} d\sigma^{N,r} = \frac{Z_N(r)}{\gamma^{\otimes N}(r)}.$$

We give a sharp estimate on the asymptotic behavior of Z'_N as $N \rightarrow \infty$.

Theorem 4.9. Consider $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$, $p \in (1, \infty]$, satisfying

$$(4.8) \quad \int_{\mathbb{R}} f v dv = 0,$$

and define

$$(4.9) \quad E := \int_{\mathbb{R}} f |v|^2 dv, \quad \Sigma := \left(\int_{\mathbb{R}} (v^2 - E)^2 f(v) dv \right)^{1/2}.$$

Then $Z_N(r), Z'_N(r)$ are well defined for all $r > 0$ and there holds with the above notations

$$(4.10) \quad Z'_N(r) \alpha_N(r^2) = \frac{\sqrt{2}}{\Sigma} \alpha_N(N) \left(\exp \left\{ - \left(\frac{r^2 - N E}{\sqrt{N} \Sigma} \right)^2 \right\} + \frac{R_N(r)}{\sqrt{N}} \right)$$

where

$$\alpha_N(s) = s^{\frac{N}{2}-1} e^{-\frac{s}{2}} \quad \text{and} \quad \|R_N\|_\infty \leq C(p, \|f\|_p, M_6(f))$$

As a particular case, there holds

$$(4.11) \quad Z'_N := Z'_N(\sqrt{E N}) = \frac{\sqrt{2}}{\Sigma} \left(1 + \mathcal{O}(N^{-\frac{1}{2}}) \right).$$

PROOF OF THEOREM 4.9. We follow the proof of [17, Theorem 14] but using the sharper estimate proved in Theorem 4.6 (instead of [17, Theorem 27]).

Before going on, let us remark that it is not obvious that $Z_N(f; r)$ is well defined for all $r > 0$ under our assumption on f which is not necessarily continuous, since we are restricting $f^{\otimes N}$ to surfaces of \mathbb{R}^N . But, in fact the product structure of $f^{\otimes N}$ makes it possible. To see this, take f and g two measurable functions equal almost everywhere, and call \mathcal{N} the negligible set on which they differ. Then the tensor products $f^{\otimes N}$ and $g^{\otimes N}$ differs only on the negligible set $\tilde{\mathcal{N}} = \cup_i \mathbb{R}^{\otimes(i-1)} \times \mathcal{N} \times \mathbb{R}^{\otimes(N-i)}$. It is not difficult to see that because of the particular structure of $\tilde{\mathcal{N}}$, the intersection of $\tilde{\mathcal{N}} \cap S_r^{N-1}$ is also σ_r^N -negligible for all $r > 0$. Therefore $f^{\otimes N}$ and $g^{\otimes N}$ are equal σ_r^N -almost everywhere on S_r^{N-1} , and there is no ambiguity in the definition of $Z_N(f, r)$ for all $r > 0$.

We now define the law g of v^2 under f

$$(4.12) \quad h(u) := \frac{1}{2\sqrt{u}} (f(\sqrt{u}) + f(-\sqrt{u})) \mathbf{1}_{u>0},$$

remarking that $h \in \mathbf{P}_3(\mathbb{R}) \cap L^q(\mathbb{R})$ with $q > 1$ as it has been shown in the proof of [17, Theorem 14]. Consider (\mathcal{V}_j) a sequence of random variables which is i.i.d. according to f . On the one hand, the law $s_N(du)$ of the random variable

$$S_N := \sum_{j=1}^N |\mathcal{V}_j|^2$$

can be computed by writing

$$\begin{aligned} \mathbb{E}(\varphi(S_N)) &= \int_0^\infty \varphi(r^2) |S_1^{N-1}| r^{N-1} \left(\int_{S_r^{N-1}} f^{\otimes N}(V) \sigma^{N,r}(dV) \right) dr \\ &= \int_0^\infty \varphi(u) |S_1^{N-1}| u^{\frac{N-1}{2}} \left(\int_{S_{\sqrt{u}}^{N-1}} f^{\otimes N}(V) \sigma^{N,\sqrt{u}}(dV) \right) \frac{du}{2\sqrt{u}}, \end{aligned}$$

which implies

$$s_N(du) = \frac{1}{2} |S_1^{N-1}| u^{\frac{N}{2}-1} Z_N(\sqrt{u}).$$

On the other hand, we have $s_N = h^{(*N)}$. Gathering these two identities, we get

$$\begin{aligned} h^{(*N)}(r^2) &= \frac{1}{2} |S_1^{N-1}| r^{N-2} Z_N(r) = \frac{\pi^{N/2}}{\Gamma(N/2)} r^{N-2} Z'_N(r) \frac{e^{-r^2/2}}{(2\pi)^{N/2}} \\ (4.13) \quad &= \frac{\alpha_N(r^2)}{\Gamma(N/2)} \frac{Z'_N(r)}{2^{N/2}}. \end{aligned}$$

Let us define $g(u) := \Sigma h(E + \Sigma u)$, so that $g \in \mathbf{P}_3(\mathbb{R}) \cap L^q(\mathbb{R})$ and

$$\int_{\mathbb{R}} g(y) y \, dy = 0, \quad \int_{\mathbb{R}} g(y) |y|^2 \, dy = 1.$$

Applying Theorem 4.6 to g and using the identity $g^{(*N)}(u) = \Sigma h^{(*N)}(N E + \Sigma u)$, we obtain

$$(4.14) \quad \sup_{r \geq 0} \left| h^{(*N)}(r^2) - \frac{1}{\sqrt{N} \Sigma} \gamma \left(\frac{r^2 - N E}{\sqrt{N} \Sigma} \right) \right| \leq \frac{C_{BE}}{N \Sigma},$$

where C_{BE} is the constant given in Theorem 4.6 and associated to g . Gathering the Stirling formula

$$(4.15) \quad \Gamma(N/2) = \sqrt{\pi N} \alpha_N(N) 2^{-\frac{N}{2}+1} \left(1 + \mathcal{O}(N^{-1/2}) \right),$$

with (4.13), (4.14), we obtain

$$\forall r > 0 \quad \left| \frac{\alpha_N(r^2) Z'_N(r)}{\sqrt{\pi N} \alpha_N(N) 2 (1 + \mathcal{O}(N^{-1/2}))} - \frac{1}{\sqrt{N} \Sigma \sqrt{2\pi}} \exp \left(\left(\frac{r^2 - N E}{\sqrt{N} \Sigma} \right)^2 / 2 \right) \right| \leq \frac{C_{BE}}{N \Sigma}.$$

Estimate (4.10) readily follows. \square

For a given $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$, $p > 1$, we define the corresponding sequence of “conditioned product measures” (according to the Kac’s spheres \mathcal{KS}_N), we write $F^N := [f^{\otimes N}]_{\mathcal{KS}_N}$, by

$$(4.16) \quad F^N := \frac{1}{Z_N(f; \sqrt{N})} f^{\otimes N} \sigma^N.$$

We show that (F^N) is well defined for N large enough and is f -chaotic.

Theorem 4.10. *Consider $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$, $p > 1$, satisfying*

$$(4.17) \quad \int_{\mathbb{R}} f v \, dv = 0 \quad \text{and} \quad \int_{\mathbb{R}} f v^2 \, dv = 1.$$

The sequence (F^N) of corresponding conditioned product measure is f -chaotic, more precisely

$$\Omega_\ell(F^N, f) := W_1(F_\ell^N, f^{\otimes \ell}) \leq \frac{1}{2} \|F_\ell^N - f^{\otimes \ell}\|_1 \leq \frac{C \ell^2}{\sqrt{N}},$$

for some constant $C = C(f) \in (0, \infty)$.

Remark 4.11. *The f -Kac’s chaoticity property of the sequence $F^N = [f^{\otimes N}]_{\mathcal{KS}_N}$ is stated and proved for smooth densities f in the seminal article by M. Kac [41]. Next, the same chaoticity property is proved with large generality (on f) in [17]. Theorem 4.10 is a “quantified” version of [17, Theorems 4 & 9] and [41, paragraph 5].*

PROOF OF THEOREM 4.10. As in Theorem 4.9, it is not obvious that F^N is well defined under our assumption on f which is not necessarily continuous, since we are restricting $f^{\otimes N}$ to a surface of \mathbb{R}^N . But the argument given at the beginning of the proof of Theorem 4.9 shows in fact that the restriction of $f^{\otimes N}$ to \mathcal{KS}_N is unambiguously defined. Since, Theorem 4.9 implies that $Z_N(f, \sqrt{N})$ is finite and non zero for N large enough, we deduce that F^N is well defined for N large enough.

Let us fix $\ell \geq 1$ and $N \geq \ell + 1$. Denoting $V = (V_\ell, V_{\ell,N})$, with $V_\ell = (v_j)_{1 \leq j \leq \ell}$, $V_{\ell,N} = (v_j)_{\ell+1 \leq j \leq N}$, we write thanks to the equality (iii) of Lemma 4.2

$$F^N(dV) = \left(\frac{f}{\gamma}\right)^{\otimes \ell}(V_\ell) \frac{1}{Z'_N(\sqrt{N})} \left(\frac{f}{\gamma}\right)^{\otimes N-\ell}(V_{\ell,N}) \sigma^{N-\ell, \sqrt{N-|V|^2}}(dV_{\ell,N}) \sigma_\ell^N(V_\ell) dV_\ell,$$

so that, coming back to the notation $V = V_\ell = (v_j)_{1 \leq j \leq \ell} \in \mathbb{R}^\ell$, we have

$$F_\ell^N(V) = \left(\prod_{j=1}^{\ell} \frac{f(v_j)}{\gamma(v_j)}\right) \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \sigma_\ell^N(V) = \left(\prod_{j=1}^{\ell} f(v_j)\right) \theta_{N,\ell}(V),$$

if we define the quantity $\theta_{N,\ell}$ by

$$(4.18) \quad \theta_{N,\ell}(V) := (2\pi)^{\frac{\ell}{2}} e^{\frac{|V|^2}{2}} \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \sigma_\ell^N(V).$$

The key point is now to prove that $\theta_{N,\ell}$ goes to 1. Recalling the Stirling formula $\Gamma(k) = \sqrt{\frac{2\pi}{k}} \left(\frac{k}{e}\right)^k (1 + \mathcal{O}(k^{-1}))$, we write σ_ℓ^N as

$$\begin{aligned} \sigma_\ell^N(V) &= \frac{|S_1^{N-\ell-1}|}{|S_1^{N-1}|} \frac{(N-|V|^2)_+^{\frac{N-\ell-2}{2}}}{N^{\frac{N-2}{2}}} \\ &= \frac{\alpha_{N-\ell}(N-|V|^2)}{N^{-\frac{1}{2}}\alpha_N(N)} \frac{e^{-\frac{|V|^2}{2}}}{(2\pi)^{\frac{1}{2}}} \mathbf{1}_{|V| \leq \sqrt{N}} (1 + \mathcal{O}(\frac{\ell^2}{N})), \end{aligned}$$

from which we deduce

$$\begin{aligned} \theta_{N,\ell}(V) &= \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \frac{\alpha_{N-\ell}(N-|V|^2)}{N^{-\frac{1}{2}}\alpha_N(N)} \mathbf{1}_{|V| \leq \sqrt{N}} (1 + \mathcal{O}(\frac{\ell^2}{N})) \\ &= \frac{\alpha_{N-\ell}(N-\ell)}{N^{-\frac{1}{2}}\alpha_N(N)} \frac{e^{-\left(\frac{\ell-|V|^2}{\sqrt{N-\ell}}\right)^2} + \mathcal{O}((N-\ell)^{-1/2})}{1 + \mathcal{O}(N^{-1/2})} \mathbf{1}_{|V| \leq \sqrt{N}} (1 + \mathcal{O}(\frac{\ell^2}{N})) \\ (4.19) \quad &= \underbrace{\left(e^{-\left(\frac{\ell-|V|^2}{\sqrt{N-\ell}}\right)^2} + \mathcal{O}((N-\ell)^{-1/2}) \right)}_{\theta_{N,\ell}^1(V)} \underbrace{\left((1 + \mathcal{O}(\frac{\ell^2}{N})) \mathbf{1}_{|V| \leq \sqrt{N}} \right)}_{\theta_{N,\ell}^2} \end{aligned}$$

where we have successively used (4.10), (4.11) the definition of $\alpha_{N-\ell}(N-\ell)$, and a calculation yielding

$$\frac{\alpha_{N-\ell}(N-\ell)}{N^{-\frac{1}{2}}\alpha_N(N)} = 1 + \mathcal{O}(\ell^2/N).$$

It implies in particular the two following estimates on $\theta_{N,\ell}$ which will also be very useful in the proof of the next theorems

$$(4.20) \quad \theta_{N,\ell}(v) \leq C \mathbf{1}_{|V| \leq \sqrt{N}}, \quad |\theta_{N,\ell}(V) - 1| \leq \frac{C\ell^2}{N^{1/2}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V| \geq N^{1/8}}.$$

Once they are proven, the conclusion follows since from the second one

$$\begin{aligned} \|F_\ell^N - f^{\otimes \ell}\|_1 &= \|(\theta_{N,\ell} - 1) f^{\otimes \ell}\|_1 \\ &\leq \frac{C\ell^2}{N} \|f\|_1 + \frac{C}{N^{1/2}} \|v^6 f\|_1. \end{aligned}$$

It only remains to prove the estimates (4.20). The first uniform estimate in (4.20) is clear from 4.19 since $\|\theta_{N,\ell}^1\|_\infty$ and $\theta_{N,\ell}^2$ are also uniformly bounded. For the second estimate, we first control

$$\begin{aligned} |\theta_{N,\ell}^1(V) - 1| &= |\theta_{N,\ell}^1(V) - 1| \mathbf{1}_{|V| \leq N^{1/8}} + |\theta_{N,\ell}^1(V) - 1| \mathbf{1}_{|V| \geq N^{1/8}} \\ &\leq |2 \left(\frac{\ell - |V|^2}{\sqrt{N - \ell \Sigma}} \right)^2 + \mathcal{O}(N^{-1/2})| \mathbf{1}_{|V| \leq N^{1/8}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V| \geq N^{1/8}} \\ &\leq \frac{C\ell^2}{N^{1/2}} \mathbf{1}_{|V| \leq N^{1/8}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V| \geq N^{1/8}}, \end{aligned}$$

which implies a similar bound for $\theta_{N,\ell}$ since

$$\begin{aligned} |\theta_{N,\ell}(V) - 1| &\leq |\theta_{N,\ell}^2(V)| |\theta_{N,\ell}^1(V) - 1| + |\theta_{N,\ell}^2(V) - 1| \\ &\leq C |\theta_{N,\ell}^1(V) - 1| + C \frac{\ell^2}{N} \\ &\leq \frac{C\ell^2}{N^{1/2}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V| \geq N^{1/8}}. \end{aligned}$$

This concludes the proof. \square

PROOF OF (1.10) IN THEOREM 1.5. The proof of the two last estimates in (1.10) follows from Theorem 4.10 together with (2.18) and (2.19). \square

4.3. Improved chaos for conditioned tensor products on the Kac's spheres. In this section, we aim to prove rate of chaoticity for stronger notions of chaos for the sequence (F^N) defined in the preceding section. Let us first recall the notion of *entropy chaos* and *Fisher information chaos* in the context of the “Kac's spheres” as they have been yet defined in the introduction. For $f \in \mathbf{P}(E)$ smooth enough, we define the usual relative entropy and usual relative Fisher information

$$H(f|\gamma) := \int_E u \log u \gamma dv, \quad I(f|\gamma) := \int_E \frac{|\nabla u|^2}{u} \gamma dv, \quad u := f/\gamma,$$

and similarly for $G^N \in \mathbf{P}_{sym}(\mathcal{KS}_N)$, we define the (normalized) relative entropy and relative Fisher information

$$H(G^N|\sigma^N) := \frac{1}{N} \int_{\mathcal{KS}_N} g^N \log g^N d\sigma^N, \quad I(G^N|\sigma^N) := \frac{1}{N} \int_{\mathcal{KS}_N} \frac{|\nabla g^N|^2}{g^N} d\sigma^N,$$

where $g^N := \frac{dG^N}{d\sigma^N}$ stands for the Radon-Nikodym derivative of G^N with respect to σ^N .

Definition 4.12. We say that a sequence (G^N) of $\mathbf{P}(\mathcal{KS}_N)$ is

i) *f-entropy chaotic* if $G_1^N \rightharpoonup f$ and

$$H(G^N|\sigma^N) \rightarrow H(f|\gamma),$$

ii) *f-Fisher information chaotic* if $G_1^N \rightharpoonup f$ and

$$I(G^N|\sigma^N) \rightarrow I(f|\gamma).$$

It is worth emphasizing again that our definition is slightly different (weaker) than the corresponding definition in [17]. But they are in fact equivalent as we shall see in next section (Theorem 4.19).

Theorem 4.13. *For any $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$, $p > 1$, satisfying the moment assumptions (4.17) of Theorem 4.10, the corresponding conditioned product sequence of measures (F^N) defined by (4.16) is f -entropy chaotic. More precisely, there exists $C = C(p, \|f\|_{L^p}, M_6(f))$ such that*

$$(4.21) \quad |H(F^N|\sigma^N) - H(f|\gamma)| \leq \frac{C}{\sqrt{N}}.$$

PROOF OF THEOREM 4.13. With the notation $F^N := [f^{\otimes N}]_{\mathcal{K}_{S_N}}$, we write for any $N \geq 1$

$$\begin{aligned} H(F^N|\sigma^N) &= \frac{1}{N} \int_{\mathcal{K}_{S_N}} \left(\log \frac{f^{\otimes N}}{Z'_N(f) \gamma^{\otimes N}} \right) dF^N \\ &= \int_{\mathbb{R}} \left(\log \frac{f}{\gamma} \right) F_1^N - \frac{1}{N} \log Z'_N(f). \end{aligned}$$

Thanks to the bound (4.10) on $Z'_N(f)$ which implies that $(Z'_N(f))$ is bounded, we deduce

$$H(F^N|\sigma^N) = \int_{\mathbb{R}} F_1^N \left(\log \frac{f}{\gamma} \right) + \mathcal{O}(1/N).$$

Recalling the notation $\theta_N := \theta_{N,1}$ defined in (4.18) and the estimates (4.20) it satisfies, we may then write

$$H(F^N|\sigma^N) = H(f|\gamma) + \underbrace{\int_{\mathbb{R}} (\theta_N - 1) f \left(\log \frac{f}{\gamma} \right)}_{=:T} + \mathcal{O}(1/N),$$

with

$$|T| \leq \underbrace{C \int_{\mathbb{R}} |\theta_N - 1| f (1 + |v|^2) dv}_{=:T_1} + \underbrace{\int_{\mathbb{R}} |\theta_N - 1| f |\log f| dv}_{=:T_2}.$$

In order to deal with T_1 , we use the second estimate of (4.20) and get

$$T_1 \leq \frac{C}{N^{1/2}} \int_{\mathbb{R}^d} \langle v \rangle^2 f dv + \frac{C}{N^{1/2}} \int_{\mathbb{R}^d} \langle v \rangle^6 f dv = \frac{C}{N^{1/2}}.$$

In order to deal with T_2 , we make the more sophisticated (but standard) splitting: for any $N, R, M \geq 1$, we write

$$\begin{aligned} T_2 &\leq \int_{B_R} |\theta_N - 1| f |\log f| + C_\theta \int_{B_R^c} f |\log f| \\ &\leq \sup_{B_R} |\theta_N - 1| C_f + C_\theta \int_{B_R^c} f (\log f)_+ \mathbf{1}_{f \geq M} + C_\theta \int_{B_R^c} f (\log f)_+ \mathbf{1}_{M \geq f \geq 1} \\ &\quad + C_\theta \int_{B_R^c} f (\log f)_- \mathbf{1}_{1 \geq f \geq e^{-|v|^2}} + C_\theta \int_{B_R^c} f (\log f)_- \mathbf{1}_{e^{-|v|^2} \geq f \geq 0}. \end{aligned}$$

For the second term, we write $f (\log f)_+ \leq f^{(1+p)/2} \leq f^p / M^{(p-1)/2}$ on $\{f \geq M\}$. For the third term, we write $f (\log f)_+ \leq f \log M \leq f (\log M) |v|^6 / R^6$ on $\{f \leq M, |v| \geq R\}$. For the fourth term, we write $\log f \geq -|v|^2$ on $\{f \geq \exp(-|v|^2)\}$, and thus $f (\log f)_- \leq f |v|^2 \leq f |v|^6 / R^4$ on $\{1 \geq f \geq e^{-|v|^2}, |v| \geq R\}$. For the last term, we write $f (\log f)_- \leq 4\sqrt{f}$ on

$\{0 \leq f \leq 1\}$, and thus $f(\log f)_- \leq 4e^{-|v|^2/2}$ on $\{e^{-|v|^2} \geq f \geq 0, |v| \geq R\}$. We deduce

$$\begin{aligned} T_2 &\leq C_f \sup_{B_R} |\theta_N - 1| + C_\theta \left(\frac{1}{M^{(p-1)/2}} + \frac{(\log M)_+}{R^6} + \frac{1}{R^4} + e^{-R} \right) \\ &\leq \frac{C(\|f\|_p, M_6(f))}{N^{1/2}}, \end{aligned}$$

with the choice $R = N^{1/8}$ (which allows to use the second estimate of (4.20)), and then $M^{(p-1)/2} = R^6$. \square

Before stating a similar result with the Fisher information, we introduce a notation: the gradient on the Kac's spheres \mathcal{KS}_N will be denoted by ∇_σ

$$\nabla_\sigma F(V) := P_{V^\perp} \nabla F(V) = \left(Id - \frac{V \otimes V}{|V|^2} \right) \nabla F(V) = \nabla F(V) - \frac{V \cdot \nabla F(V)}{N} V,$$

if F is a smooth function on \mathbb{R}^N . P_{V^\perp} stands for the projection on the hyperplan perpendicular to V . We will use many times that

$$(4.22) \quad \nabla \left[F \left(\frac{V}{|V|} \right) \right] = \frac{1}{|V|} P_{V^\perp} \nabla F \left(\frac{V}{|V|} \right) = \frac{1}{|V|} \nabla_\sigma F \left(\frac{V}{|V|} \right).$$

Theorem 4.14. *For any $f \in \mathbf{P}_6(\mathbb{R})$, satisfying the moment assumptions (4.17) of Theorem 4.10, the corresponding conditioned product sequence of measures (F^N) defined by (4.16) satisfies*

$$\sup_{N \in \mathbb{N}} I(F^N | \sigma_N) < +\infty$$

if $I(f) < +\infty$. If moreover

$$\int_{\mathbb{R}} \frac{f'(v)^2}{f(v)} \langle v \rangle^2 dv < +\infty,$$

the sequence F^N is Fisher information chaotic.

PROOF OF THEOREM 4.14. We only proof the second point. The first point (boundness of the Fisher information) can be deduced from the above proof. It suffices in fact to use the simple bound $|\nabla_\sigma G| \leq |\nabla G|$ instead of equality (4.23).

Remark also that the bound on the Fisher information implies that f is continuous and uniformly bounded since $E = \mathbb{R}$. Therefore, the L^p (for $p > 1$) assumption which is necessary in theorem 4.10 is implied by our bound on the Fisher information. We can therefore apply the estimates (4.20) on the quantity $\theta_{N,i}$ for $i = 1, 2$ defined in (4.18). They imply in particular that $\|\theta_{N,i}\|_\infty$ is uniformly bounded and that $\theta_{N,i}$ converges point-wise towards 1. We start with the formula

$$I(F^N | \sigma^N) = \frac{1}{N} \int_{\mathcal{KS}_N} |\nabla_\sigma \ln \left(\frac{f^{\otimes N}}{\gamma^{\otimes N}} \right)|^2 F^N(dV).$$

As ∇_σ is the projection on the Kac's spheres of the usual gradient, we have from (4.22) for any function G on \mathbb{R}^N

$$(4.23) \quad |\nabla_\sigma G(V)|^2 = |\nabla G(V)|^2 - \frac{1}{N} |V \cdot \nabla G(V)|^2.$$

Using this with $G = \ln \left(\frac{f^{\otimes N}}{\gamma^{\otimes N}} \right)$ in the Fisher information formula, it comes

$$I(F^N | \sigma^N) = \frac{1}{N} \int_{\mathcal{KS}_N} \left| \nabla \ln \left(\frac{f^{\otimes N}}{\gamma^{\otimes N}} \right) \right|^2 F^N(dV) - \frac{1}{N^2} \int_{\mathcal{KS}_N} \left| V \cdot \nabla \ln \left(\frac{f^{\otimes N}}{\gamma^{\otimes N}} \right) \right|^2 F^N(dV).$$

Recalling that $F_1^N = f \theta_{N,1}$ from (4.18), by symmetry, the first term in the right hand side is equal to

$$(4.24) \quad \int_{\mathbb{R}} \left| \partial_v \ln \frac{f(v)}{\gamma(v)} \right|^2 F_1^N(dv) = I(f|\gamma) + \int_{\mathbb{R}} \left| \frac{\nabla f(v)}{f(v)} + v \right|^2 (\theta_{N,1}(v) - 1) f(v) dv.$$

The last term goes to zero from the hypothesis on f , the uniform bound $|\theta_{N,1}| \leq C$ and the pointwise convergence of $\theta_{N,1}$ to 1. To handle the second term in the RHS of (4.24), we compute

$$\begin{aligned} \frac{1}{N^2} \left| V \cdot \nabla \ln \left(\frac{f}{\gamma} \right)^{\otimes N} \right|^2 &= \frac{1}{N^2} \left(\sum_{i=1}^N v_i \left[\ln \frac{f}{\gamma} \right]'(v_i) \right)^2 \\ &= \frac{1}{N^2} \sum_{i=1}^N v_i^2 \left(\left[\ln \frac{f}{\gamma} \right]'(v_i) \right)^2 + \frac{1}{N^2} \sum_{i \neq j}^N v_i v_j \left[\ln \frac{f}{\gamma} \right]'(v_i) \left[\ln \frac{f}{\gamma} \right]'(v_j). \end{aligned}$$

After integration, it comes thanks to the symmetry of F^N

$$\begin{aligned} \frac{1}{N^2} \int_{\mathcal{K}_{S_N}} \left| V \cdot \nabla \ln \left(\frac{f}{\gamma} \right)^{\otimes N} \right|^2 F^N(dV) &= \frac{1}{N} \int_{\mathbb{R}} v^2 \left(\left[\ln \frac{f}{\gamma} \right]'(v) \right)^2 F_1^N(dv) \\ &\quad + \frac{N-1}{N} \int_{\mathbb{R}^2} v_1 v_2 \left[\ln \frac{f}{\gamma} \right]'(v_1) \left[\ln \frac{f}{\gamma} \right]'(v_2) F_2^N(dv_1, dv_2). \end{aligned}$$

Using the uniform bound $F_1^N(v) = \theta_{N,1}(v) f(v) \leq C f(v)$, and the hypothesis on f , we obtain that the first term of the r.h.s. is bounded by $\frac{C}{N}$. The second term denoted by $R_2(N)$ is equal to

$$\begin{aligned} R_2(N) &= \frac{N-1}{N} \int_{\mathbb{R}^2} v_1 v_2 \left[\ln \frac{f}{\gamma} \right]'(v_1) \left[\ln \frac{f}{\gamma} \right]'(v_2) f(v_1) f(v_2) dv_1 dv_2 \\ &\quad + \frac{N-1}{N} \int_{\mathbb{R}^2} v_1 v_2 \left[\ln \frac{f}{\gamma} \right]'(v_1) \left[\ln \frac{f}{\gamma} \right]'(v_2) (\theta_{N,2}(v_1, v_2) - 1) f(v_1) f(v_2) dv_1 dv_2 \\ &= \frac{N-1}{N} \left(\int_{\mathbb{R}} (v f'(v) + v^2 f(v)) dv \right)^2 + R_3(N) = R_3(N), \end{aligned}$$

after an integration by parts and because of the equality $\int v^2 f(dv) = 1$. The term $R_3(N)$ goes to zero by dominated convergence since

$$\begin{aligned} \int_{\mathbb{R}^2} v_1 v_2 \left| \left[\ln \frac{f}{\gamma} \right]'(v_1) \left[\ln \frac{f}{\gamma} \right]'(v_2) \right| f(v_1) f(v_2) dv_1 dv_2 &= \left(\int_{\mathbb{R}} v \left| \left[\ln \frac{f}{\gamma} \right]'(v) \right| f(v) dv \right)^2 \\ &\leq I(f|\gamma) \int_{\mathbb{R}} v^2 df. \end{aligned}$$

This concludes the proof. \square

4.4. Chaos for arbitrary sequence of probability measures on the Kac's spheres.
In that last section, we aim to present the relationship between Kac's chaos, entropy chaos and Fisher information chaos in the Kac's spheres framework.

We begin with a result which is the analogous for probability measures on the Kac's spheres to the lower semi continuity of the Entropy and Fisher information yet established on product spaces.

Theorem 4.15. *For any sequence (G^N) of $\mathbf{P}(\mathcal{KS}_N)$ such that $G_j^N \rightharpoonup G_j$ weakly in $\mathbf{P}(E^j)$, there holds*

$$H(G_j|\gamma^{\otimes j}) \leq \liminf H(G^N|\sigma^N), \quad I(G_j|\gamma^{\otimes j}) \leq \liminf I(G^N|\sigma^N).$$

For the proof, we shall need the following integration by parts formula on the Kac' spheres, which proof is postponed to the end the proof of Theorem 4.15 .

Lemma 4.16. *Assume that F (resp. Φ) is a function (resp. vector field in \mathbb{R}^N) on the Kac's spheres \mathcal{KS}_N with integrable gradient. Then the following integration by part formula holds*

$$(4.25) \quad \int_{\mathcal{KS}_N} \left[\nabla_\sigma F(V) \cdot \Phi(V) + F(V) \operatorname{div}_\sigma \Phi(V) - \frac{N-1}{N} F(V) \Phi(V) \cdot V \right] d\sigma^N(V) = 0$$

where $\operatorname{div}_\sigma$ stands for the divergence on the sphere, given by

$$\operatorname{div}_\sigma \Phi(V) := \sum_{i=1}^N \nabla_\sigma \Phi_i(V) \cdot e_i = \operatorname{div} \Phi(V) - \sum_{i=1}^N \frac{V \cdot \nabla \Phi_i(V)}{|V|^2} v_i$$

where the last formula is useful only if Φ is defined on a neighborhood of the sphere.

PROOF OF THEOREM 4.15. We refer to [17, Theorem 17] for a proof of the inequality involving the entropy and we give only the proof of the second inequality, which in fact relies on the characterization $I^{(3)}$ of the Fisher information. Precisely, the previous Lemma 4.16 can be used to get a reformulation of the Fisher information relative to σ^N on the sphere

$$(4.26) \quad \begin{aligned} I_N(G^N|\sigma^N) &:= \int_{\mathcal{KS}_N} |\nabla_\sigma \ln G^N|^2 G^N(dV) = \sup_{\Phi \in C_b^1(\mathbb{R}^N)^N} \int_{\mathcal{KS}_N} \left(\nabla \ln G^N \cdot \Phi - \frac{|\Phi|^2}{4} \right) G^N(dV) \\ &= \sup_{\Phi \in C_b^1(\mathbb{R}^N)^N} \int_{\mathcal{KS}_N} \left(\frac{N-1}{N} \Phi(V) \cdot V - \operatorname{div}_\sigma \Phi(V) - \frac{|\Phi(V)|^2}{4} \right) G^N(dV). \end{aligned}$$

Next applying the equality (3.13) to the probability measure $\gamma^{\otimes j}$, we get that for or any $\varepsilon > 0$, we can choose a $\varphi \in C_b^1(\mathbb{R}^j)^j$ such that

$$\frac{1}{j} I_j(F^j|\gamma^{\otimes j}) - \varepsilon \leq \frac{1}{j} \int_{\mathbb{R}^j} \left(\varphi \cdot V_j - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) F^j(dV_j).$$

Remark that the r.h.s. is quite similar to (4.26). With the notation $N = nj + r$, $0 \leq r < j$ and $V_N = (V_{j,1}, \dots, V_{j,n}, V_r)$, we define

$$\Phi(V_N) := (\varphi(V_{j,1}, \dots, \varphi(V_{j,n}), 0) \in C_b^1(\mathbb{R}^N)^N,$$

and use it in the equality (4.26). We get

$$\begin{aligned} \frac{1}{N} I(G^N|\sigma^N) &\geq \frac{1}{N} \int_{\mathcal{KS}_N} \left(\frac{N-1}{N} \Phi(V_N) \cdot V_N - \operatorname{div}_\sigma \Phi(V_N) - \frac{|\Phi(V_N)|^2}{4} \right) G^N(dV_N) \\ &\geq \frac{n}{N} \int_{\mathbb{R}^j} \left(\frac{N-1}{N} \varphi(V_j) \cdot V_j - \operatorname{div} \varphi(V_j) - \frac{|\varphi(V_j)|^2}{4} \right) G_j^N(dV_j) + \frac{R_\varphi(N)}{N}, \end{aligned}$$

where

$$\begin{aligned} R_\varphi(N) &= \frac{1}{N} \int \left(\sum_{i=1}^N [V \cdot \nabla \Phi_i(V)] v_i \right) G^N(dV_N) = \frac{1}{N} \int \left(\sum_{i,\ell}^N \frac{\partial \Phi_i}{\partial v_\ell} v_i v_\ell \right) G^N(dV_N) \\ &= \frac{n}{N} \int \left(\sum_{i,\ell}^j \frac{\partial \varphi_i}{\partial v_\ell} v_i v_\ell \right) G_j^N(dV_j) = O(1), \end{aligned}$$

if $\nabla \varphi$ decrease sufficiently quickly at infinity. Passing to the limit, we get

$$\liminf_{N \rightarrow +\infty} I(G^N | \sigma^N) \geq \frac{1}{j} \int_{\mathbb{R}^j} \left(\varphi \cdot V_j - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) F^j(dV_j) \geq I(F^j | \gamma^{\otimes j}) - \varepsilon$$

which concludes the proof. \square

PROOF OF LEMMA 4.16 As before, we will use the normalized norm $|V|_2 := \sqrt{\frac{1}{N} \sum v_i^2}$. Choosing any smooth function q on $(0, +\infty)$ with compact support, we define

$$w(V) := q(|V|_2) F\left(\frac{V}{|V|_2}\right) \Phi\left(\frac{V}{|V|_2}\right).$$

Its divergence is given by

$$\begin{aligned} \operatorname{div} w &= \frac{q'(|V|_2)}{N} F\left(\frac{V}{|V|_2}\right) \Phi\left(\frac{V}{|V|_2}\right) \cdot \frac{V}{|V|_2} + \frac{q(|V|_2)}{|V|_2} \nabla_\sigma F\left(\frac{V}{|V|_2}\right) \Phi\left(\frac{V}{|V|_2}\right) \cdot \frac{V}{|V|_2} \\ &\quad + \frac{q(|V|_2)}{|V|_2} F\left(\frac{V}{|V|_2}\right) \operatorname{div}_\sigma \Phi\left(\frac{V}{|V|_2}\right). \end{aligned}$$

Integrating this equality, and using polar coordinate, we get

$$\begin{aligned} 0 &= \left(\int_{\mathcal{KS}_N} [\nabla_\sigma F(V) \cdot \Phi(V) + F(V) \operatorname{div}_\sigma \Phi(V)] \sigma^N(dV) \right) \left(\int_0^\infty q(r) r^{N-2} dr \right) \\ &\quad + \frac{1}{N} \left(\int_{\mathcal{KS}_N} F(V) \Phi(V) \cdot V \sigma^N(dV) \right) \left(\int_0^\infty q'(r) r^{N-1} dr \right). \end{aligned}$$

Since $\int_0^\infty q'(r) r^{N-1} dr = -(N-1) \int_0^\infty q(r) r^{N-2} dr$, we obtain

$$\int_{\mathcal{KS}_N} \left[\nabla_\sigma F(V) \cdot \Phi(V) + F(V) \operatorname{div}_\sigma \Phi(V) - \frac{N-1}{N} F(V) \cdot \Phi(V) \cdot V \right] d\sigma^N(V) = 0,$$

which is the claimed result. \square

The next theorem will be the key estimate in the proof of the variant of Theorem 1.4 adapted to the Kac's spheres. It relies on the HWI inequality on the Kac's spheres, which allows to quantify the convergence of the relative entropy.

Theorem 4.17. *Consider (G^N) a sequence of $\mathbf{P}(\mathcal{KS}_N)$ which is f -chaotic, $f \in \mathbf{P}(E)$. Assume furthermore that*

$$M_k(G^N)^{\frac{1}{k}} \leq K \text{ for } k \geq 6, \quad \text{and} \quad I(G^N | \sigma^N) \leq K.$$

Then f satisfies $M_k(f) < \infty$, $I(f) < \infty$, and (G^N) is f -entropy chaotic. More precisely, there exists $C_1 := C_1(K)$ and for any $\gamma_2 < \frac{1}{8} \frac{k-2}{k+1}$ a constant $C_2(\gamma_2)$ such that

$$|H(G^N | \sigma^N) - H(f | \gamma)| \leq C_1 \left(W_1(G^N, f^{\otimes N})^{\gamma_1} + C_2 N^{-\gamma_2} \right),$$

with $\gamma_1 := 1/2 - 1/k$.

The proof uses the following estimate

Theorem 4.18. ([18, Theorem 2], [5, Theorem 2]). *For any sequence (G^N) of $\mathbf{P}(\mathcal{KS}_N)$, there hold*

$$\forall 1 \leq k \leq N, \quad H(G_k^N | \sigma_k^N) \leq 2 H(G^N | \sigma^N) \quad \text{and} \quad I(G_k^N | \sigma_k^N) \leq 2 I(G^N | \sigma^N).$$

PROOF OF THEOREM 4.17. STEP 1. Thanks to Theorem 4.18, we have

$$I(G_1^N | \sigma_1^N) \leq 2K.$$

Using the strong convergence of σ_1^N to γ stated in 4.2, we pass to the (inferior) limit and get

$$I(f | \gamma) \leq 2K \quad \text{and then} \quad I(f) \leq 2K.$$

Introducing the restriction $F^N = f^{\otimes N} / Z(\sqrt{N})\sigma^N$ of $f^{\otimes N}$ to \mathcal{KS}_N defined in (4.16) and using point i) of Theorem 4.14, we get

$$\sup_N I(F^N | \sigma^N) \leq C_2.$$

Step 2. Because the Ricci curvature of the metric space \mathcal{KS}_N is positive (it is $K := (N - 1)/N$) we may use the HWI inequality in weak $CD(K, \infty)$ geodesic space (see [73, Theorem 30.21]) which generalizes the standard HWI inequality (3.16) quoted in Proposition 3.8. However, we have to be careful, because it is now valid with \tilde{W}_2 replaced by the MKW distance constructed with the geodesic distance on the sphere, and not with the distance induced by the square norm of \mathbb{R}^N . Fortunately, both distances are equivalent, and if we add a constant $\frac{\pi}{2}$ in the right hand side, we can still write the HWI inequality with our usual distance W_2 . We then have

$$H(F^N | \sigma^N) - H(G^N | \sigma^N) \leq \frac{\pi}{2} \sqrt{I(F^N | \sigma^N)} W_2(F^N, G^N),$$

and

$$H(G^N | \sigma^N) - H(F^N | \sigma^N) \leq \frac{\pi}{2} \sqrt{I(G^N | \sigma^N)} W_2(F^N, G^N),$$

so that

$$|H(F^N | \sigma^N) - H(G^N | \sigma^N)| \leq C_2 W_2(F^N, G^N).$$

We rewrite it under the form

$$|H(G^N | \sigma^N) - H(f | \gamma)| \leq C_3 [W_2(G^N, f^{\otimes N}) + W_2(F^N, f^{\otimes N})] + |H(F^N | \sigma^N) - H(f | \gamma)|.$$

For the first term, we have using inequality of Lemma 2.2

$$W_2(G^N, f^{\otimes N}) \leq 4K W_1(G^N, f^{\otimes N})^{1/2-1/k}.$$

For the second term, we have for any $\varepsilon > 0$

$$\begin{aligned} W_2(F^N, f^{\otimes N}) &\leq 4K \Omega_N(F^N; f)^{1/2-1/k} \\ &\leq 4K \left(\Omega_\infty(F^N; f) + C_\varepsilon N^{-\frac{1}{2+\varepsilon+2/k}} \right)^{1/2-1/k} \\ &\leq C_\varepsilon \left(\Omega_2(F^N; f)^{\frac{1}{2+\varepsilon+1/k}} + C_\varepsilon N^{-\frac{1}{2+\varepsilon+2/k}} \right)^{1/2-1/k} \\ &\leq C_\varepsilon N^{-\frac{1/4-1/2k}{2+\varepsilon+2/k}}, \end{aligned}$$

where we have successively used Lemma 2.2, the inequality (2.18), (2.19) and Theorem 4.10 in the case $d = 1$ (and then $d' = \max(d, 2) = 2$). The third and last term is bounded by $C N^{-1/2}$ thanks to Theorem 4.13. \square

The lower semi continuity properties of Theorem 4.15 and Theorem 4.17 allow us to give a variant of Theorem 1.4 in the framework of probability measures with support on the Kac's spheres.

Theorem 4.19. *Consider (G^N) a sequence of $\mathbf{P}_{\text{sym}}(\mathcal{KS}_N)$ such that $M_6(G_1^N)$ is bounded and $G_1^N \rightharpoonup f$ weakly in $\mathbf{P}(\mathbb{R})$.*

In the list of assertions below, each one implies the assertion which follows:

- (i) (G^N) is f -Fisher information chaotic, i.e. $I(G^N|\sigma^N) \rightarrow I(f|\gamma)$, $I(f) < \infty$;
- (ii) (G^N) is f -Kac's chaotic and $I(G^N|\sigma^N)$ is bounded;
- (iii) (G^N) is f -entropy chaotic, that is $H(G^N|\sigma^N) \rightarrow H(f|\gamma)$, $H(f) < \infty$;
- (iv) (G^N) is f -Kac's chaotic.

PROOF OF THEOREM 4.19. The proof is very similar to the one of Theorem 1.4. $i) \Leftrightarrow ii)$ and $iii) \Leftrightarrow iv)$ relies on the l.s.c. properties of Theorem 4.15. And $ii) \Leftrightarrow iii)$ uses Theorem 4.17. We omit the details. \square

We finally conclude this section with the proof of Theorem 1.6.

PROOF OF THEOREM 1.6. We only deal with the case $j = 1$, but the general case $j \geq 1$ can be managed in a very similar way because we already know that $G_j^N \rightharpoonup f^{\otimes j}$ weakly in $\mathbf{P}(E^j)$ thanks to Theorem 4.17 and Theorem 4.19. With the notations of Theorem 1.6, we have to prove

$$H(G_1^N|f) = \int_E \log(G_1^N/f) G_1^N \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

First, we observe that since G^N is symmetric and has support on the Kac's spheres, $M_2(G^N) = 1$. Moreover,

$$\begin{aligned} I(G_1^N|\sigma_1^N) &= \int_E |\nabla \log G_1^N - \nabla \log \sigma_1^N|^2 G_1^N \\ &= I(G_1^N) + \int_E [2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2] G_1^N, \end{aligned}$$

so that

$$I(G_1^N) \leq I(G_1^N|\sigma_1^N) + \int_E (2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2)_- G_1^N.$$

We easily compute

$$\begin{aligned} 2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2 &= \\ &= \frac{N-3}{2} \left\{ 2 \frac{(2v)^2/N^2}{(1-v^2/N)^2} - 2 \frac{2/N}{(1-v^2/N)} + \frac{(2v/N)^2}{(1-v^2/N)^2} \right\} \mathbf{1}_{v^2 \leq N} \end{aligned}$$

and then

$$\begin{aligned} (2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2)_- &= 2 \frac{N-3}{N} \frac{(4v^2/N-1)_-}{(1-v^2/N)^2} \mathbf{1}_{v^2 \leq N/4} \\ &\leq 2 \frac{1}{(1-1/4)^2} = \frac{32}{9}. \end{aligned}$$

Thanks to the boundedness assumption (1.12) we get that $I(G_1^N) \leq C$ for some constant $C \in (0, \infty)$, and then $I(G_1^N|\gamma) \leq 2[I(G_1^N) + M_2(G_1^N)] \leq C$.

Next, we introduce the splitting

$$H(G_1^N | f) = \underbrace{H(G_1^N | \gamma) - H(f | \gamma)}_{=: T_1} + \underbrace{\int_E (f - G_1^N) \log \frac{f}{\gamma}}_{=: T_2}$$

and we show that $T_i \rightarrow 0$ for any $i = 1, 2$. For the first term T_1 , using twice the HWI inequality we have

$$|T_1| \leq \left(\sqrt{I(G_1^N | \gamma)} + \sqrt{I(f | \gamma)} \right) W_2(G_1^N, f) \rightarrow 0$$

because of the uniform bound on the Fisher information and of the convergence property $W_2(G_1^N, f) \rightarrow 0$. That last convergence is a consequence of [72, Theorem 7.2 (iii) \Rightarrow (i)], $G_1^N \rightharpoonup f$ weakly when $N \rightarrow \infty$ and $\langle G_1^N, v^2 \rangle = \langle f, v^2 \rangle$ for any $N \geq 1$ when $k = 2$, and it is a consequence of [72, Theorem 7.2 (ii) \Rightarrow (i)], $G_1^N \rightharpoonup f$ weakly as $N \rightarrow \infty$ and $M_k(G_1^N) \leq C$ for any $N \geq 1$ when $k > 2$.

Before dealing with the last term, we remark that the bound on the Fisher information of f implies some regularity, precisely that \sqrt{f} and then f are $\frac{1}{2}$ -Hölder. Therefore $\ln \frac{f}{\gamma}$ is continuous and satisfies from the assumption (1.13) the bound

$$\left| \ln \frac{f}{\gamma} \right| \leq \ln \|f\|_\infty + \alpha |v|^{k'} + |\beta| + \frac{v^2}{2} \leq C \langle v \rangle^{\max(k', 2)}.$$

We then conclude that $T_2 \rightarrow 0$ by using [72, Theorem 7.2 (iii) \Rightarrow (iv)] when $k = 2$ and [72, Theorem 7.2 (ii) \Rightarrow (iv)] when $k > 2$. \square

5. ON MIXTURES ACCORDING TO DE FINETTI, HEWITT AND SAVAGE

In this section we develop a quantitative and qualitative approach concerning the sequence of probability measures of $\mathbf{P}_{sym}(E^N)$, $E \subset \mathbb{R}^d$, in the general framework of convergence to “mixture of probability measures” (here we do not assume chaos property).

Depending on the result, we will need some hypothesis on the set E that we will make precise in each statement. While in the first and second sections the results hold with great generality only assuming that

- E is a Borel set of \mathbb{R}^d ;

we shall assume in the third and fourth sections that

- $E = \mathbb{R}^d$ or E is an open set of \mathbb{R}^d with smooth boundary in order that the strong maximum principle and the Hopf lemma hold (that we furthermore assume to be bounded in the third section);

and we shall also assume in the fourth section that

- the normalized non relative HWI inequality (3.15) holds in E (e.g. it satisfies the assumptions of Proposition 3.8).

5.1. The De Finetti, Hewitt and Savage theorem and weak convergence in $\mathbf{P}(E^N)$. We begin by recalling the famous De Finetti, Hewitt and Savage theorem [24, 40] for which we state a quantified version that is maybe new.

Theorem 5.1. *Assume $E \subset \mathbb{R}^d$ is a Borel set. Consider a sequence (π^j) of symmetric and compatible probability measures of $\mathbf{P}(E^j)$, that is $\pi^j \in \mathbf{P}_{sym}(E^j)$ and $(\pi^j)_{|E^\ell} = \pi^\ell$ for any $1 \leq \ell \leq j$, and consider $(\hat{\pi}^j)$ the associated sequence of empirical distribution*

in $\mathbf{P}(\mathbf{P}(E))$ defined according to (2.7). For any $s > \frac{d}{2}$, the sequence $(\hat{\pi}^j)$ is a Cauchy sequence for the distance $\mathcal{W}_{H^{-s}}$, and precisely

$$(5.1) \quad [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 \leq 2\|\Phi_s\|_\infty \left(\frac{1}{M} + \frac{1}{N} \right),$$

where Φ_s is the function introduced in Lemma 2.9. In particular, the sequence $(\hat{\pi}^j)$ converges towards some $\pi \in \mathbf{P}(\mathbf{P}(E))$ with the speed $\mathcal{W}_{H^{-s}}(\hat{\pi}^j, \pi) \leq \frac{C}{\sqrt{j}}$. The limit π is characterized by the relations

$$(5.2) \quad \forall j \geq 1, \quad \pi^j = \pi_j := \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \quad \text{in } \mathbf{P}_{\text{sym}}(E^j),$$

or in other words, with the notations of section 2.1

$$(5.3) \quad \forall \varphi \in C_b(E^j) \quad \langle \pi^j, \varphi \rangle = \int_{\mathbf{P}(E)} R_\varphi(\rho) \pi(d\rho).$$

Reciprocally, for any mixture of probability measures $\pi \in \mathbf{P}(\mathbf{P}(E))$, the sequence (π_j) of probability measures in $\mathbf{P}(E^j)$ defined by the second identity in (5.2) is such that the π_j are symmetric and compatible.

PROOF OF THEOREM 5.1. We split the proof into two steps.

Step 1. In order to estimate the distance between $\hat{\pi}^N$ and $\hat{\pi}^M$ we shall use as in the proof of Proposition 2.10 the fact that $\|\cdot\|_{H^{-s}}^2$ is a polynomial on $\mathbf{P}(E)$, but we have to choose a good transference plan. Fortunately, there is at least one simple choice. The compatibility and symmetry conditions on (π^N) tell us that π^{N+M} is an admissible transference between π^N and π^M . Using the symmetry of π^{N+M} and the isometry between $(E^N/\mathfrak{S}_N, w_1)$ and $(\mathbf{P}_N(E), W_1)$ stated in step 1 in the proof of Proposition 2.14, we will interpret it as a transference plan $\tilde{\pi}^{N+M}$ on $\mathcal{P}_N(E) \times \mathcal{P}_M(E)$ between $\hat{\pi}^N$ and $\hat{\pi}^M$. More precisely, $\tilde{\pi}^{N+M} \in \mathbf{P}(\mathbf{P}(E) \times \mathbf{P}(E))$ is defined as the probability measure satisfying

$$\forall \Phi \in C_b(\mathbf{P}(E) \times \mathbf{P}(E)) \quad \langle \tilde{\pi}^{N+M}, \Phi \rangle = \int_{E^N \times E^M} \Phi(\mu_X^N, \mu_Y^N) \pi^{N+M}(dX, dY).$$

With that transference plane we have

$$\begin{aligned} [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho - \eta\|_{H^{-s}}^2 \tilde{\pi}^{N+M}(d\rho, d\eta) \\ &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \left(\int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\rho^{\otimes 2} - \rho \otimes \eta) \right. \\ &\quad \left. + (\eta^{\otimes 2} - \eta \otimes \rho)](dx, dy) \right) \tilde{\pi}^{N+M}(d\rho, d\eta), \end{aligned}$$

with the help of (2.25). We can then compute

$$\begin{aligned}
& [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 \leq \\
& \leq \int \left(\int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\mu_X^N)^{\otimes 2} - \mu_X^N \otimes \mu_Y^M](dx, dy) \right) \pi^{N+M}(dX, dY) \\
& \quad + \int \left(\int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\mu_Y^M)^{\otimes 2} - \mu_Y^M \otimes \mu_X^N](dx, dy) \right) \pi^{N+M}(dX, dY) \\
& \leq \int \left(\frac{1}{N^2} \sum_{i,j=1}^N \Phi_s(x_i - x_j) - \frac{1}{NM} \sum_{i,j'=1}^M \Phi_s(x_i - y_{j'}) \right) \pi^{N+M}(dX, dY) \\
& \quad + \int \left(\frac{1}{M^2} \sum_{i',j'=1}^M \Phi_s(y_{i'} - y_{j'}) - \frac{1}{NM} \sum_{i',j=1}^M \Phi_s(x_{i'} - y_j) \right) \pi^{N+M}(dX, dY) \\
& \leq \frac{\Phi_s(0)}{N} + \frac{N-1}{N} \int \Phi_s(x-y) \pi^2(dx, dy) - \int \Phi_s(x-y) \pi^2(dx, dy) \\
& \quad + \frac{\Phi_s(0)}{M} + \frac{M-1}{M} \int \Phi_s(x-y) \pi^2(dx, dy) - \int \Phi_s(x-y) \pi^2(dx, dy),
\end{aligned}$$

and we conclude with

$$\begin{aligned}
[\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 & \leq \left(\frac{1}{M} + \frac{1}{N} \right) \left(\Phi_s(0) - \int \Phi_s(x-y) \pi^2(dx, dy) \right) \\
& \leq 2\|\Phi_s\|_\infty \left(\frac{1}{M} + \frac{1}{N} \right).
\end{aligned}$$

The existence of the limit π is due to the completeness of $\mathbf{P}(\mathbf{P}(E))$.

Step 2. Now it remains to characterize the limit π . We fix $j \in \mathbb{N}$, we denote by π_j its j -th marginal defined thanks to the second identity in (5.2) and by $\hat{\pi}_j^N = (\hat{\pi}^N)_j$ the j -th marginal of the empirical probability measure $\hat{\pi}^N$ as defined in (2.9). We easily compute

$$\begin{aligned}
\|\hat{\pi}_j^N - \pi_j\|_{H^{-s}}^2 & = \left\| \int_{\mathbf{P}(E)} \rho^{\otimes j} \hat{\pi}^N(d\rho) - \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \right\|_{H^{-s}}^2 \\
& = \inf_{\Pi \in \Pi(\hat{\pi}^N, \pi)} \left\| \int_{\mathbf{P}(E)} [\rho^{\otimes j} - \eta^{\otimes j}] \Pi(d\rho, d\eta) \right\|_{H^{-s}}^2 \\
& \leq \inf_{\Pi \in \Pi(\hat{\pi}^N, \pi)} \int_{\mathbf{P}(E)} \|\rho^{\otimes j} - \eta^{\otimes j}\|_{H^{-s}}^2 \Pi(d\rho, d\eta) \\
& = [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \pi)]^2 \leq \frac{C}{N}.
\end{aligned}$$

Next we fix $s > \frac{jd}{2}$, so that using Sobolev embeddings on \mathbb{R}^{jd} , $\|\varphi\|_\infty \leq C\|\varphi\|_{H^s}$ for any $\varphi \in H^s(\mathbb{R}^{jd})$, which implies by duality that $\|\rho\|_{H^{-s}} \leq C\|\rho\|_{TV}$ for any $\rho \in \mathbf{P}(\mathbb{R}^{jd})$. Using the Grunbaum lemma 2.8 and the compatibility assumption $\pi_j^N = \pi_j$, we get the inequality

$$\|\pi_j^N - \hat{\pi}_j^N\|_{H^{-s}} = \|\pi_j^N - \hat{\pi}_j^N\|_{H^{-s}} \leq C\|\pi_j^N - \hat{\pi}_j^N\|_{TV} \leq \frac{Cj^2}{N}.$$

Combining the two previous inequalities leads to

$$\|\pi^j - \pi_j\|_{H^{-s}} \leq \|\pi^j - \hat{\pi}_j^N\|_{H^{-s}} + \|\hat{\pi}_j^N - \pi_j\|_{H^{-s}} \leq \frac{C}{\sqrt{N}} + \frac{Cj^2}{N},$$

which implies the claimed equality in the limit $N \rightarrow +\infty$. \square

Let us now introduce some definitions. For $k > 0$, we define

$$\mathbf{P}_k(\mathbf{P}(E)) := \{\pi \in \mathbf{P}(\mathbf{P}(E)); M_k(\pi) := M_k(\pi_1) < \infty\}$$

and for $k, a > 0$, we define

$$\mathcal{BP}_{k,a}(E^N) := \{F \in \mathbf{P}(E^N); M_k(F_1) \leq a\}.$$

Definition 5.2. For given sequences $(F^N)_N$ of $\mathbf{P}_{sym}(E^N)$, $(\pi_n)_n$ of $\mathbf{P}(\mathbf{P}(E))$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$, we say that

- (F^N) is bounded in $\mathbf{P}_k(E^N)$ if there exists $a > 0$ such that $M_k(F_1^N) \leq a$;
- (π_n) is bounded in $\mathbf{P}_k(\mathbf{P}(E))$ if there exists $a > 0$ such that $M_k(\pi_{n,1}) \leq a$;
- (F^N) weakly converges to π in $\mathbf{P}_k(E^j)_{\forall j}$, we write $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}_k(E^j)_{\forall j}$, if (F^N) is bounded in $\mathbf{P}_k(E^N)$ and $F_j^N \rightharpoonup \pi_j$ weakly in $\mathbf{P}(E^j)$ for any $j \geq 1$;
- (π_n) weakly converges to π in $\mathbf{P}_k(\mathbf{P}(E))$ if (π_n) is bounded in $\mathbf{P}_k(\mathbf{P}(E))$ and $\pi_n \rightharpoonup \pi$ weakly in $\mathbf{P}(\mathbf{P}(E))$.

With that (not conventional) definitions, any bounded sequence in $\mathbf{P}_k(\mathbf{P}(E))$ is weakly compact in $\mathbf{P}_k(\mathbf{P}(E))$, and for any sequence (F^N) of probability measures of $\mathbf{P}_{sym}(E^N)$ which is bounded in $\mathbf{P}_k(E^N)$, $k > 0$, there exists a subsequence $(F^{N'})$ and a mixture of probability measures $\pi \in \mathbf{P}_k(\mathbf{P}(E))$ such that $F^{N'} \rightharpoonup \pi$ in $\mathbf{P}(E^j)_{\forall j}$.

We now present a result about the equivalence of convergences for sequence of $\mathbf{P}_{sym}(E^N)$, $N \rightarrow \infty$, without any chaos hypothesis.

Theorem 5.3. Assume $E \subset \mathbb{R}^d$ is a Borel set.

(1) Consider (F^N) a sequence of $\mathbf{P}_{sym}(E^N)$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$. The three following assertions are equivalent:

- (i) $F^N \rightharpoonup \pi$ in $\mathbf{P}(E^j)_{\forall j}$, that is $F_j^N \rightharpoonup \pi_j$ weakly in $\mathbf{P}(E^j)$ for any $j \geq 1$;
- (ii) $\hat{F}^N \rightharpoonup \pi$ weakly in $\mathbf{P}(\mathbf{P}(E))$;
- (iii) $W_1(F^N, \pi_N) \rightarrow 0$.

(2) For any $\gamma \in [\frac{1}{2d'}, \frac{1}{d'})$ (recall that $d' = \max(d, 2)$), and any $k > \frac{d'}{\gamma-1-d'} \geq 1$, there exists a constant $C = C(\gamma, d, k)$ such that the following estimate holds

$$(5.4) \quad \forall N \geq 1 \quad |W_1(F^N, \pi_N) - W_1(\hat{F}^N, \pi)| \leq \frac{C M_k(\pi_1)^{1/k}}{N^\gamma}.$$

(3) With the same notations as in the second point, we have for any mixture of probability measures $\alpha, \beta \in \mathbf{P}(\mathbf{P}(E))$

$$(5.5) \quad W_1(\hat{\alpha}_j, \alpha) \leq \frac{C M_k(\alpha_1)^{1/k}}{j^\gamma},$$

where $\hat{\alpha}_j$ is empirical probability distribution in $\mathbf{P}(\mathbf{P}(E))$ associated to the j -th marginal $\alpha_j \in \mathbf{P}(E^j)$, as well as

$$(5.6) \quad W_1(\alpha, \beta) - \frac{C(M_k(\alpha_1)^{\frac{1}{k}} + M_k(\beta_1)^{\frac{1}{k}})}{j^\gamma} \leq W_1(\alpha_j, \beta_j) \leq W_1(\alpha, \beta).$$

PROOF OF THEOREM 5.3. *Step 1.* Equivalence between (i) and (ii) is classical. Let us just sketch the proof. For any $\varphi \in C_b(E^j)$ we have from the Grunbaum lemma recalled in Lemma 2.8 that

$$\begin{aligned}\langle \hat{F}^N, R_\varphi \rangle &= \langle F^N, \varphi \otimes \widetilde{\mathbf{1}^{\otimes N-j}} \rangle + \mathcal{O}(j^2/N) \\ &= \langle F_j^N, \varphi \rangle + \mathcal{O}(j^2/N).\end{aligned}$$

We deduce that the convergence $\langle \hat{F}^N, R_\varphi \rangle \rightarrow \langle \pi, R_\varphi \rangle$ is equivalent to the convergence $\langle F_j^N, \varphi \rangle \rightarrow \langle \pi_j, \varphi \rangle$ since that $\langle \pi, R_\varphi \rangle = \langle \pi_j, \varphi \rangle$ thanks to Theorem 5.1.

Therefore, *i)* is equivalent to the convergence $\langle \hat{F}^N, \Phi \rangle \rightarrow \langle \pi, \Phi \rangle$ for any polynomial function $\Phi \in C_b(\mathbf{P}(E))$. But now, the family of probability measures \hat{F}^N (and π) belongs to the compact subset of $\mathbf{P}(\mathbf{P}(E))$

$$\mathcal{K} := \{\alpha \in \mathbf{P}(\mathbf{P}(E)), \text{ s.t. } \alpha_1 = F_1\},$$

and also any converging subsequence $\hat{F}^{N'}$ should converge weakly towards a probability measure $\tilde{\pi}$ having the same marginals as π . Since by Theorem 5.1 marginals uniquely characterize a probability measure on $\mathbf{P}(\mathbf{P}(E))$, it implies $\tilde{\pi} = \pi$ and then weak convergence against polynomial function implies the standard weak convergence of probability measures *ii)*.

It is classical that the MKW distance is a metrization of the weak convergence of measures. Even in that "abstract" case, (ii) is equivalent to $\mathcal{W}_1(\hat{F}^N, \pi) \rightarrow 0$ (recall that the distance chosen in order to define \mathcal{W}_1 is bounded). Thus, for sequences having a bounded moment $M_k(F_1^N)$ for some $k > 0$, the equivalence between (ii) and (iii) will be a consequence of (5.4). For sequences for which no moment M_k is bounded, the same conclusion is true. The correct argument still relies on a version of inequality (5.4), with a slower and less explicit rate of convergence, which can be obtained from an adaptation of Lemma 2.1.

Step 2. We now prove (5.4). For $\hat{\pi}_N$ we have the following representation:

$$(5.7) \quad \hat{\pi}_N = \int \widehat{\rho^{\otimes N}} \pi(d\rho) = \int \widehat{\rho^{\otimes N}} \pi(d\rho).$$

Thanks to Proposition 2.14, we may compute

$$\begin{aligned}|W_1(F^N, \pi_N) - W_1(\hat{F}^N, \pi)| &= |W_1(\hat{F}^N, \hat{\pi}_N) - W_1(\hat{F}^N, \pi)| \\ &\leq W_1(\hat{\pi}_N, \pi) = W_1\left(\int_{\mathbf{P}(E)} \widehat{\rho^{\otimes N}} \pi(d\rho), \int_{\mathbf{P}(E)} \delta_\rho \pi(d\rho)\right) \\ &\leq \int_{\mathbf{P}(E)} W_1(\widehat{\rho^{\otimes N}}, \delta_\rho) \pi(d\rho) = \int_{\mathbf{P}(E)} \Omega_\infty(\rho) \pi(d\rho), \\ &\leq \frac{C(d, \gamma, k)}{N^\gamma} \int_{\mathbf{P}(E)} M_k(\rho)^{1/k} \pi(d\rho) \leq \frac{C(d, \gamma, k)}{N^\gamma} M_k(\pi_1)^{1/k},\end{aligned}$$

where we have successively used the triangular inequality for the \mathcal{W}_1 distance, the relation (5.7), the convexity property of the \mathcal{W}_1 distance and the definition of the chaos measure Ω_∞ . We also used the bound (2.30) and the Jensen inequality (recall that $1/k \in (0, 1]$) in the last line.

Step 3. We now prove the third point. For the first inequality, choose $s = \frac{1}{2\gamma} - \frac{d}{2k}$. Then by our assumptions, $s > \max(1, \frac{d}{2})$ and we can apply Lemma 2.3 on the comparison of

distances in $\mathbf{P}(\mathbf{P}(E))$ and Theorem 5.1 to get

$$\mathcal{W}_1(\hat{\alpha}_j, \alpha) \leq C M_k(\alpha_1)^{\frac{1}{k}} \mathcal{W}_{H-s}(\hat{\alpha}_j, \alpha)^{\frac{2k}{d+2ks}} \leq \frac{C M_k(\alpha_1)^{\frac{1}{k}}}{j^\gamma}$$

For the first part of the second inequality (5.6) we write

$$\mathcal{W}_1(\alpha, \beta) \leq \mathcal{W}_1(\alpha, \hat{\alpha}_j) + \mathcal{W}_1(\hat{\alpha}_j, \hat{\beta}_j) + \mathcal{W}_1(\hat{\beta}_j, \beta),$$

we use the inequality just proved above and the identity (2.14). The second part of the second inequality (5.6) is a mere application of Lemma 2.7. \square

5.2. Level-3 Boltzmann entropy functional for mixtures. In this section we recover some well known results on the Boltzmann entropy for mixture of probability measures as stated in [2] and proved by Robinson and Ruelle in [64]. However our proof differs from the one of [64], and in particular it does not use the abstract representation result of Choquet and Meyer [22] but an abstract Lemma 5.6 that we introduce for our purposes.

Let us assume that $E \subset \mathbb{R}^d$ is a Borel set and let us fix a real number $m > 0$. Then, for any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ we define

$$(5.8) \quad \mathcal{H}(\pi) := \int_{\mathbf{P}(E)} H(\rho) \pi(d\rho),$$

where H is the Boltzmann's entropy defined on $\mathbf{P}_m(E)$.

Theorem 5.4. (1) *The functional $\mathcal{H} : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$ is proper, affine and l.s.c. with respect to the weak convergence in $\mathbf{P}_m(\mathbf{P}(E))$. Moreover, for any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$, there holds*

$$(5.9) \quad \mathcal{H}(\pi) = \sup_{j \in \mathbb{N}^*} H(\pi_j) = \lim_{j \rightarrow \infty} H(\pi_j),$$

where π_j is the j -th marginal of π defined in Theorem 5.1 and H is the normalized Boltzmann's entropy defined on $\mathbf{P}_m(E^j)$ for any $j \geq 1$.

(2) *Consider (F^N) a sequence of $\mathbf{P}_{\text{sym}}(E^N)$ and $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ such that $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}_m(E^j)_{\forall j}$. Then*

$$(5.10) \quad \mathcal{H}(\pi) \leq \liminf_{N \rightarrow \infty} H(F^N).$$

The proof of Theorem 5.4 uses the two following lemmas.

Lemma 5.5. *For any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ we define*

$$\mathcal{H}'(\pi) := \sup_{j \in \mathbb{N}^*} H(\pi_j).$$

The functional $\mathcal{H}' : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$ is affine, proper, and l.s.c. for the weak convergence, and

$$(5.11) \quad \mathcal{H}'(\pi) = \lim_{j \rightarrow \infty} H(\pi_j).$$

The proof of Lemma 5.5 is classical. For the sake of completeness we nevertheless present it.

PROOF OF LEMMA 5.5. Thanks to (3.1), for any $j \geq 1$, we have

$$H(\pi_j) \geq \log c_m - \int_E |v|^m d\pi_1$$

so that \mathcal{H}' is proper on $\mathbf{P}_m(\mathbf{P}(E))$. It is also l.s.c. as the supremum of l.s.c. functions, since H_j is l.s.c. on $\mathbf{P}_m(E^j)$ as it has been recalled in Lemma 3.1 and since the inequality of the right of (5.6) shows that $\pi \mapsto \pi_j$ is also continuous for the weak convergence of measures.

As a second step, we establish (5.11). For any fixed $\ell \geq 1$ and any $j \geq \ell$ we introduce the Euclidean decomposition $j = n\ell + r$, $0 \leq r \leq \ell - 1$, and a direct iterative application of inequality (3.7) together with (3.1) imply

$$\begin{aligned} H_j(\pi_j) &\geq n H_\ell(\pi_\ell) + H_r(\pi_r) \\ &\geq n H_\ell(\pi_\ell) + (j - 1) [(\log c_m)_- - \mathcal{M}_m(\pi)]. \end{aligned}$$

We deduce that for any $\ell \geq 1$

$$\liminf_{j \rightarrow \infty} H(\pi_j) \geq \liminf_{j \rightarrow \infty} \frac{n}{j} H_\ell(\pi_\ell) = H(\pi_\ell),$$

from which (5.11) follows.

We conclude by establishing the affine property of \mathcal{H}' . Let us consider $F, G \in \mathbf{P}_m(\mathbf{P}(E))$ and $\theta \in (0, 1)$, and let us assume that $H(F_j) < \infty$, $H(G_j) < \infty$ for any $j \geq 1$, the case when $H(F_j) = \infty$ or $H(G_j) = \infty$ being trivial. Using that $s \mapsto \log s$ is an increasing function and that $s \mapsto s \log s$ is a convex function, we have

$$\begin{aligned} H(\theta F_j + (1 - \theta) G_j) &= \frac{1}{j} \int_{E^j} (\theta F_j + (1 - \theta) G_j) \log(\theta F_j + (1 - \theta) G_j) \\ &\geq \frac{1}{j} \int_{E^j} \{\theta F_j \log(\theta F_j) + (1 - \theta) G_j \log((1 - \theta) G_j)\} \\ &= \theta H(F_j) + (1 - \theta) H(G_j) + \frac{1}{j} [\theta \log \theta + (1 - \theta) \log(1 - \theta)] \\ &\geq H(\theta F_j + (1 - \theta) G_j) + \frac{1}{j} [\theta \log \theta + (1 - \theta) \log(1 - \theta)]. \end{aligned}$$

Passing to the limit $j \rightarrow \infty$ in the two preceding inequalities and using (5.11), we get

$$\mathcal{H}'(\theta F + (1 - \theta) G) \geq \theta \mathcal{H}'(F) + (1 - \theta) \mathcal{H}'(G) \geq \mathcal{H}'(\theta F + (1 - \theta) G),$$

which is nothing but the announced affine property. \square

We establish now in the following abstract lemma the last argument which allows us to prove the first equality in (5.9) and which will be useful in the next section in order to get the same property for the similar functionals on $\mathbf{P}_m(\mathbf{P}(E))$ built starting from the Fisher information.

Lemma 5.6. *Consider a sequence (K_j) of functionals on $\mathbf{P}_m(E^j)$, $m \geq 0$, such that*

- (i) $K_j : \mathbf{P}_m(E^j) \rightarrow \mathbb{R} \cup \{+\infty\}$ *is convex, proper and l.s.c. for the weak convergence of measures on $\mathbf{P}_m(E^j)$ for any $j \geq 1$. Moreover, either $m = 0$ and K_j is positive for each j , or $m > 0$ and there exists $k \in (0, m)$, a constant $C_k \in \mathbb{R}^+$ such that the functional*

$$\mathbf{P}(E^j) \rightarrow \mathbb{R} \cup \{\infty\}, \quad G \mapsto K_j(G) + j[C_k + M_k(G)]$$

is nonnegative and is l.s.c with respect to the weak convergence in $\mathbf{P}(E)$.

- (ii) $j^{-1} K_j(f^{\otimes j}) = K_1(f)$ *for all $f \in \mathbf{P}_m(E)$ and $j \geq 1$.*
- (iii) $K_j(G) \geq K_\ell(G_\ell) + K_r(G_r)$ *for any $G \in \mathbf{P}(E^j)$ and any ℓ, r such that $j = \ell + r$.*

- (iv) The functional $\mathcal{K}' : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{+\infty\}$ defined for any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ by (this a part of the theorem that the sup equals the lim)

$$\mathcal{K}'(\pi) := \sup_{j \geq 1} \frac{1}{j} K_j(\pi_j) = \lim_{j \rightarrow +\infty} \frac{1}{j} K_j(\pi_j),$$

where π_j denotes j -th marginal defined thanks to Theorem 5.1, is affine in the following sense. For any probability measure $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ and any partition of $\mathbf{P}_m(E)$ by some sets ω_i , $1 \leq i \leq M$, such that ω_i is an open set in $E \setminus (\omega_1 \cup \dots \cup \omega_{i-1})$ for any $1 \leq i \leq M-1$, $\omega_M = \mathbf{P}_m(E) \setminus (\omega_1 \cup \dots \cup \omega_{M-1})$ and $\pi(\omega_i) > 0$ for any $1 \leq i \leq M$, defining

$$\alpha_i := \pi(\omega_i) \quad \text{and} \quad \gamma^i := \frac{1}{\alpha_i} \mathbf{1}_{\omega_i} \pi \in \mathbf{P}_m(\mathbf{P}(E))$$

so that

$$\pi = \alpha_1 \gamma^1 + \dots + \alpha_M \gamma^M \quad \text{and} \quad \alpha_1 + \dots + \alpha_M = 1,$$

there holds

$$\mathcal{K}'(\pi) = \alpha_1 \mathcal{K}'(\gamma^1) + \dots + \alpha_M \mathcal{K}'(\gamma^M).$$

Then under the above assumptions, for any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$, there holds

$$\mathcal{K}'(\pi) = \mathcal{K}(\pi) := \int_{\mathbf{P}(E)} K_1(\rho) \pi(d\rho).$$

The functional $\mathcal{K} : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$ is affine, proper and l.s.c. with respect to the weak convergence in $\mathbf{P}_m(\mathbf{P}(E))$.

Moreover, it satisfies the following Γ -l.s.c. property. For any sequence F^N of $\mathbf{P}_{sym}(E^N)$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$ such that $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}_m(E^j)_{\forall j}$, then

$$(5.12) \quad \mathcal{K}(\pi) \leq \liminf_{N \rightarrow \infty} K(F^N).$$

PROOF OF LEMMA 5.6. We split the proof into five steps.

Step 1. A fist inequality $\mathcal{K} \geq \mathcal{K}'$ We skip the proof that the lim equals the sup in point iv). This is a consequence of the hypothesis iii) - and the bound by below in point i) in the case $m > 0$ - and has already been proved in the proof of Lemma 5.5 for the entropy. We fix $\pi \in \mathbf{P}_m(\mathbf{P}(E))$. Thanks to assumptions (i) and (ii), we easily compute

$$\begin{aligned} \mathcal{K}(\pi) &= \int_{\mathbf{P}(E)} \frac{1}{j} K_j(\rho^{\otimes j}) \pi(d\rho) \\ &\geq \frac{1}{j} K_j \left(\int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \right) = \frac{1}{j} K_j(\pi_j). \end{aligned}$$

Taking the supremum over j in this inequality, we get a first inequality

$$\mathcal{K}(\pi) \geq \sup_{j \geq 1} \frac{1}{j} K_j(\pi_j) = \mathcal{K}'(\pi).$$

Step 2. \mathcal{J} is l.s.c. on $\mathbf{P}_m(E)$ with respect to the W_1 -metric.

We consider the case when $m > 0$, and choose $k \in (0, m)$ such that (i) holds. We explain in the step 3' below the necessary adaptation to do in the case $m = 0$.

For any $\delta > 0$, by compactness, we can find a family of finite cardinal \mathcal{N} of balls $\mathcal{B}_i := \mathcal{B}(\rho_i, \delta) = \{\rho \in \mathbf{P}_m(E); W_1(\rho, \rho_i) < \delta\}$, $\rho_i \in \mathcal{B}\mathbf{P}_{m,1/\delta}$, of radius δ so that

$$\mathcal{B}\mathbf{P}_{m,1/\delta} \subset \bigcup_{i=1}^{\mathcal{N}} \mathcal{B}_i.$$

We associate to that partition and "almost" partition of unity by

$$\phi_i(\rho) := 2 \left[1 - \frac{W_1(\rho, \rho_i)}{2\delta} \right]_+, \quad \theta_i(\rho) := \frac{\phi_i(\rho)}{\sum_{j=1}^{\mathcal{N}} \phi_j(\rho) + \delta}.$$

Finally, we set for any $\rho \in \mathbf{P}_m(E)$

$$J^\delta(\rho) := \sum_{i=1}^{\mathcal{N}} \theta_i(\rho) J_i^\delta,$$

where

$$J_i^\delta := \inf_{\rho \in \mathcal{B}(\rho_i, 2\delta)} J(\rho) \quad \text{and} \quad J(\rho) := K_1(\rho) + C_k + M_k(\rho).$$

We claim that by construction the functional J^δ is Lipschitz with respect to the W_1 metric on $\mathbf{P}(E)$, and satisfies

$$(5.13) \quad \forall \rho \in \mathbf{P}_m(E), \quad \frac{\mathbf{1}_{\mathcal{BP}_{m,1/\delta}}(\rho)}{1 + \delta} \inf_{\rho' \in \mathcal{B}(\rho, 4\delta)} J(\rho') \leq J^\delta(\rho) \leq J(\rho),$$

where $\mathbf{1}$ denote the indicator function. To obtain both inequalities, we introduce $I^\delta(\rho) := \{1 \leq i \leq \mathcal{N}, W_1(\rho, \rho_i^\delta) \leq 2\delta\}$, and rewrite

$$J^\delta(\rho) := \sum_{i \in I^\delta(\rho)} \theta_i(\rho) J_i^\delta.$$

But for any i such that $W_1(\rho, \rho_i) \leq 2\delta$ we have

$$\inf_{\rho' \in \mathcal{B}(\rho_i, 4\delta)} J(\rho') \leq J_i^\delta = \inf_{\rho' \in \mathcal{B}(\rho_i, 2\delta)} J(\rho') \leq J(\rho).$$

The upper bound in (5.13) follows from the second inequality (on the right). Since $J(\rho) \geq 0$ by hypothesis (i), the first above inequality implies that

$$J^\delta(\rho) \geq \left(\sum_{i \in I^\delta(\rho)} \theta_i(\rho) \right) \inf_{\rho' \in \mathcal{B}(\rho, 4\delta)} J(\rho') \geq \frac{\sum_{j=1}^{\mathcal{N}} \phi_j(\rho)}{\sum_{j=1}^{\mathcal{N}} \phi_j(\rho) + \delta} \inf_{\rho' \in \mathcal{B}(\rho, 4\delta)} J_1(\rho').$$

The bound by below in (5.13) then follows because any $\rho \in \mathcal{BP}_{m,1/\delta}$ is at least in one of the \mathcal{B}_i for some i , and then $\sum_{j=1}^{\mathcal{N}} \phi_j(\rho) \geq \phi_i(\rho) \geq 1$. The inequalities (5.13) and the hypothesis that J is l.s.c. with respect to the weak convergence on $\mathbf{P}(E)$ implies that

$$(5.14) \quad \forall \rho \in \mathbf{P}_m(E), \quad \lim_{\delta \rightarrow 0} J^\delta(\rho) = J(\rho).$$

We can now introduce the functionals \mathcal{J}^δ and \mathcal{J} defined for all $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ by

$$\begin{aligned} \mathcal{J}^\delta(\pi) &:= \int_{\mathbf{P}_m(E)} J^\delta(\rho) \pi(d\rho) \\ \mathcal{J}(\pi) &:= \int_{\mathbf{P}_m(E)} J(\rho) \pi(d\rho) = \mathcal{K}(\pi) + C_k + \mathcal{M}_k(\pi). \end{aligned}$$

Since J^δ is Lipschitz with respect to the W_1 -metric, the Kantorovich-Rubinstein duality theorem [72, Theorem 1.14] implies that the functionals \mathcal{J}^δ is continuous with respect to

the \mathcal{W}_1 -metric. Moreover, the upper bound in (5.13) implies that $\mathcal{J}^\delta(\pi) \leq \mathcal{J}(\pi)$, for any $\pi \in \mathbf{P}_m(\mathbf{P}(E))$. Finally, an application of Fatou's Lemma together with (5.14) implies

$$\begin{aligned} \liminf_{\delta \rightarrow 0} \mathcal{J}^\delta(\pi) &= \liminf \int_{\mathbf{P}_m(E)} J^\delta(\rho) \pi(d\rho) \geq \int_{\mathbf{P}_m(E)} \liminf J^\delta(\rho) \pi(d\rho) \\ &\geq \int_{\mathbf{P}_m(E)} J(\rho) \pi(d\rho) = \mathcal{J}(\pi). \end{aligned}$$

All in all, we get that

$$\forall \pi \in \mathbf{P}_m(\mathbf{P}(E)), \quad \mathcal{J}(\pi) = \sup_{\delta > 0} \mathcal{J}^\delta(\pi),$$

and that implies that \mathcal{J} is l.s.c. with respect to the \mathcal{W}_1 -metric since the \mathcal{J}^δ are continuous with respect to that metric.

Step 2'. A necessary adaptation in the case $m = 0$. In that case, things are in some sense simpler since the functional K is already positive, so that we may try directly to apply Step 2 with $J = K_1$. However, there is one difficulty : the compact sets $\mathcal{BP}_{m,1/\delta}$ does not covers $\mathbf{P}(E)$; even if we take their union for $\delta > 0$ and $m > 0$.

However, we can still do a correct proof if we fix π at the beginning. We then choose a increasing function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that

$$(5.15) \quad \lim_{v \rightarrow +\infty} g(v) = +\infty \quad \text{and} \quad M_g(\pi_1) := \int_E g(\langle v \rangle) \pi_1(dv) < \infty.$$

Then we can restrict ourselves to the set $\mathbf{P}_g := \{\rho \in \mathbf{P}(E), M_g(\rho) < +\infty\}$, since the last hypothesis on g implies that $\pi(\mathbf{P}_g(E)) = 1$. If we now replace in step 2, the sets $\mathcal{BP}_{m,1/\delta}$ by the still compact sets

$$\mathcal{BP}_{g,1/\delta} := \{\rho, M_g(\rho) \leq \delta^{-1}\},$$

and follow the same strategy, we will conclude that $\mathcal{K}(\pi) = \sup_{\delta > 0} \mathcal{K}^\delta(\pi)$ were the \mathcal{K}^δ will be continuous with respect to the \mathcal{W}_1 -metric. It implies that \mathcal{K} is l.s.c. at π . Since π is arbitrary, \mathcal{K} is globally l.s.c.

Step 3. \mathcal{K} is l.s.c. with respect to the weak convergence of measures on $\mathbf{P}_m(\mathbf{P}(E))$.

In the case $m = 0$, that step is useless since in step 2' we proved that $\mathcal{K} = \mathcal{J}$ is l.s.c.. So it remains only to treat the case $m > 0$. Since $\mathcal{J} = \mathcal{K} + \mathcal{M}_k + C_k$ is l.s.c. with respect to the \mathcal{W}_1 -metric on $\mathbf{P}_m(\mathbf{P}(E))$, the conclusion will follows if we show that \mathcal{M}_k is continuous with respect to the weak convergence on $\mathbf{P}_m(\mathbf{P}(E))$, defined in Definition 5.2.

For this, we choose $\rho, \mu \in \mathbf{P}_m(E)$. Since

$$\forall v, v' \in E \quad |\langle v \rangle^k - \langle v' \rangle^k| \leq k \min(1, |v - v'|) (\langle v \rangle^k + \langle v' \rangle^k),$$

we obtain if we chose an optimal transference plan π (for the distance d_E on E) between ρ and μ

$$\begin{aligned} |M_k(\rho) - M_k(\mu)| &\leq \int |\langle v \rangle^k - \langle v' \rangle^k| \pi(dv, dv') \\ &\leq k \int d_E(v, v') (\langle v \rangle^k + \langle v' \rangle^k) \pi(dv, dv') \\ &\leq k \left(\int d_E(v, v')^{\frac{m}{m-k}} \pi(dv, dv') \right)^{1-\frac{k}{m}} (M_m(\rho) + M_m(\mu))^{\frac{k}{m}}, \end{aligned}$$

so that

$$|M_k(\rho) - M_k(\mu)| \leq k (M_m(\rho) + M_m(\mu))^{\frac{k}{m}} W_1(\rho, \mu)^{1-\frac{k}{m}},$$

where we have used Hölder inequality and the fact that $d_E \leq 1$. Choosing now two $\alpha, \beta \in \mathbf{P}_m(\mathbf{P}(E))$ and an optimal transference plan π (for the distance W_1 on $\mathbf{P}(E)$) between them, we get

$$\begin{aligned} |\mathcal{M}_k(\alpha) - \mathcal{M}_k(\beta)| &= \left| \int M_k(\rho) - M_k(\rho') \pi(d\rho, d\rho') \right| \\ &\leq k \int (M_m(\rho) + M_m(\rho'))^{\frac{k}{m}} W_1(\rho, \rho')^{1-\frac{k}{m}} \pi(d\rho, d\rho'), \end{aligned}$$

and then

$$|\mathcal{M}_k(\alpha) - \mathcal{M}_k(\beta)| \leq k (\mathcal{M}_m(\alpha) + \mathcal{M}_m(\beta))^{\frac{k}{m}} \mathcal{W}_1(\alpha, \beta)^{1-\frac{k}{m}},$$

where we have used Hölder inequality. This concludes the step since weak convergence on $\mathbf{P}_m(\mathbf{P}(E))$ exactly means that \mathcal{W}_1 goes to zero and the moment of order m are bounded.

Step 4. Proof of the remaining inequality $\mathcal{K}' \geq \mathcal{K}$. Because $\mathbf{P}_m(E)$ endowed with the MKW distance W_1 is a Polish space, for any fixed $\varepsilon > 0$, we can cover it by a countable union of balls $\mathcal{B}_n := \mathcal{B}(f_n, \varepsilon)$ of radius ε . For a given $\pi \in \mathbf{P}_m(\mathbf{P}(E))$, we can choose M such that

$$\omega_M := \mathbf{P}_m(E) \setminus (\mathcal{B}_1 \cup \dots \cup \mathcal{B}_{M-1}) \quad \text{satisfies} \quad \pi(\omega_M) \leq \varepsilon$$

and denote $\omega_i := \mathcal{B}_i \setminus (\mathcal{B}_1 \cup \dots \cup \mathcal{B}_{i-1})$ for all $1 \leq i \leq M-1$. We define then

$$\alpha_i := \pi(\omega_i), \quad \gamma_i := \frac{1}{\alpha_i} \pi|_{\omega_i}, \quad \pi^M := \sum_{i=1}^M \alpha_i \delta_{\gamma_i^1}, \quad \gamma_1^i = \int_{\mathbf{P}(E)} \rho \gamma^i(d\rho).$$

For any $1 \leq i \leq M$, we have

$$\mathcal{K}'(\gamma^i) := \sup_{j \geq 1} \frac{1}{j} K_j(\gamma_j^i) \geq K_1(\gamma_1^i).$$

Using the affine property (iv) of \mathcal{K}' , the above inequality and the definitions of π^M and \mathcal{K} , we get

$$\begin{aligned} \mathcal{K}'(\pi) &= \alpha_1 \mathcal{K}'(\gamma^1) + \dots + \alpha_M \mathcal{K}'(\gamma^M) \\ (5.16) \quad \mathcal{K}'(\pi) &\geq \alpha_1 K_1(\gamma_1^1) + \dots + \alpha_M K_1(\gamma_1^M) = \mathcal{K}(\pi^M). \end{aligned}$$

We observe that because $\pi_1^M = \pi_1$, we have

$$\langle \pi_1^M, |v|^m \rangle = \langle \pi_1, |v|^m \rangle = M_m(\pi) < \infty,$$

and in particular $\pi^M \in \mathbf{P}_m(\mathbf{P}(E))$. Moreover, defining $T^M : \mathbf{P}(E) \rightarrow \{\gamma^1, \dots, \gamma^M\}$ by $T^M(\rho) = \gamma^i$ for any $\rho \in \omega_i$, we have $\pi^M = (T^M)_\# \pi$ and then

$$\mathcal{W}_1(\pi, \pi^M) \leq \langle (id \otimes T^M)_\# \pi, W_1(., .) \rangle \leq 2\varepsilon.$$

We consider now a sequence $\varepsilon \rightarrow 0$ and the corresponding sequence (π^M) for which we then have by construction $\pi^M \rightharpoonup \pi$ weakly in $\mathbf{P}_m(\mathbf{P}(E))$. Inequality (5.16), the above convergence and the l.s.c. property of \mathcal{K} proved in step 2 and 3 imply the second (and reverse) inequality

$$\mathcal{K}(\pi) \leq \liminf_{M \rightarrow \infty} \mathcal{K}(\pi^M) \leq \mathcal{K}'(\pi).$$

Step 5. The Γ -l.s.c. property of \mathcal{K} . We give the proof only in the case $m > 0$, the case $m = 0$ being simpler. We consider (F^N) a sequence of $\mathbf{P}_{sym}(E^N)$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$ such that $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}_m(E^j)_{\forall j}$, in particular $M_m(F_1^N) \leq a$ for some $a \in (0, \infty)$. For any fixed $j \geq 1$, using the l.s.c. property of K_j , introducing the Euclidean decomposition

$N = nj + r$, $0 \leq r \leq j - 1$ and using iteratively the inequality (iii) of the hypothesis as in the proof of Lemma 5.5 as well as the lower bound on K_r provided by hypothesis (i), we get

$$\begin{aligned}
\frac{1}{j} K_j(\pi_j) &\leq \liminf_{N \rightarrow \infty} \frac{1}{j} K_j(F_j^N) \\
&\leq \liminf_{N \rightarrow \infty} \frac{1}{nj} \{K_N(F^N) - r K_r(F_r^N)\} \\
&\leq \liminf_{N \rightarrow \infty} \left\{ \frac{1}{nj} K_N(F^N) + \frac{1}{n} (C_m + a) \right\} \\
&= \liminf_{N \rightarrow \infty} N^{-1} K_N(F^N).
\end{aligned}$$

We deduce (5.10) thanks to (5.9). That concludes the proof. \square

PROOF OF THEOREM 5.4. The proof is just an application of the two previously proved lemmas. First, let us observe that H_j , \mathcal{H} and \mathcal{H}' fulfill the assumptions of Lemma 5.6 since that (i) is nothing but (3.5), (ii) is a consequence of Lemma 3.1, (iii) is nothing but (3.7), and a stronger version of (iv) has been established in Lemma 5.5. Then (5.9) and (5.10) are exactly the conclusion of Lemma 5.6 applied to the entropy. \square

5.3. Level-3 Fisher information for mixtures. We state now a similar result for the Fisher information for mixtures of probability measures.

Let us assume that $E = \mathbb{R}^d$ or E is an open connected and bounded set of \mathbb{R}^d with smooth boundary. Then, for any $\pi \in \mathbf{P}(\mathbf{P}(E))$ we define

$$(5.17) \quad \mathcal{I}(\pi) := \int_{P(E)} I(\rho) \pi(d\rho),$$

where I is the Fisher information defined on $\mathbf{P}(E)$.

Theorem 5.7. (1) *The functional $\mathcal{I} : \mathbf{P}(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$ is affine, nonnegative and l.s.c. for the weak convergence. Moreover, for any $\pi \in \mathbf{P}(\mathbf{P}(E))$, there holds*

$$(5.18) \quad \mathcal{I}(\pi) = \sup_{j \in \mathbb{N}^*} I(\pi_j) = \lim_{j \rightarrow \infty} I(\pi_j),$$

where I stands for the normalized Fisher information defined in $\mathbf{P}(E^j)$ for any $j \geq 1$.

(2) *Consider (F^N) a sequence of $\mathbf{P}_{\text{sym}}(E^N)$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$ such that $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}(E^j)_{\forall j}$. Then*

$$(5.19) \quad \mathcal{I}(\pi) \leq \liminf I(F^N).$$

As for Theorem 5.4, the proof of Theorem 5.7 relies on the abstract lemma 5.6. The hypothesis of that lemma are proved to be true in the lemma 5.10 below. Two useful intermediate results are stated in the next two lemmas.

Lemma 5.8. *There exist :*

- a family of regularizing operators $S_t : \mathbf{P}(E) \rightarrow \mathbf{P}(E)$ defined for any $t > 0$,
- a family (C_t) of positive constants
- a family ε_t of positive constants such that $\varepsilon_t \rightarrow 0$ when $t \rightarrow 0$
- for any $k > 0$, a family (ε'_{kt}) of positive constants so that $\varepsilon'_{kt} \rightarrow 0$ when $t \rightarrow 0$ such that for any $\rho \in \mathbf{P}(E)$ and any $t > 0$, denoting $\rho_t := S_t(\rho)$ we have

$$(5.20) \quad I(\rho_t) \leq I(\rho), \quad M_k(\rho_t) \leq 2^k (M_k(\rho) + \varepsilon_{kt}), \quad \|\nabla \ln \rho_t\|_\infty \leq C_t$$

and $W_1(\rho, \rho_t) \leq \varepsilon_t$.

PROOF OF LEMMA 5.8. We only consider the case $E = \mathbb{R}^d$. The case when E is a smooth bounded open set can be handled similarly by using for (ρ_t) the solution of the heat equation (with Neumann boundary conditions) and the strong maximum principle. We define

$$\eta_t(z) := \frac{C_d}{t^d} e^{-\langle z/t \rangle} = \frac{C_d}{t^d} e^{-\sqrt{1+(|z|/t)^2}} \quad \text{and} \quad \rho_t := \eta_t * \rho.$$

Observing that

$$\frac{|\nabla \eta_t(z)|}{\eta_t(z)} = \frac{1}{t} \frac{|z|}{\langle z \rangle} \leq \frac{1}{t},$$

we deduce that for any $x \in \mathbb{R}^d$, we have

$$|\nabla \rho_t(x)| \leq \frac{1}{t} \int_{\mathbb{R}^d} \eta_t(x-y) \rho(y) dy = \frac{1}{t} \rho_t.$$

The inequality on the moment of order k is a consequence of the inequality

$$\langle x+y \rangle^k \leq 2^k (\langle x \rangle^k + \langle y \rangle^k),$$

which leads to the claimed inequality with $\varepsilon_{kt} = M_k(\eta_t) = t^k M_k(\eta_1)$.

As ρ_t is also an average of translations of ρ (which has the same Fisher information as ρ), the convexity of the Fisher information implies that

$$I(\rho_t) = I\left(\int \rho(\cdot - z) \eta_t(dz)\right) \leq \int I(\rho(\cdot - z)) \eta_t(dz) = I(\rho).$$

We finally observe that for any $\rho \in \mathbf{P}(E)$ there holds

$$W_1(\rho, \rho_t) = W_1(\rho, \rho * \eta_t) \leq \int_{\mathbb{R}^d} |z| \eta_t(z) dz = C_d t,$$

and that proves the last estimate. \square

Lemma 5.9. Consider $\pi \in \mathbf{P}(\mathbf{P}(E))$ and define the regularized family $\pi_t \in \mathbf{P}(\mathbf{P}(E))$, for $t > 0$, by push-forward by S_t , $\pi_t := S_{t\#}\pi$ or equivalently

$$\langle \pi_t, \Phi \rangle = \langle \pi, \Phi_t \rangle \quad \forall \Phi \in C_b(\mathbf{P}(E))$$

where $\Phi_t \in C_b(\mathbf{P}(E))$ is defined by $\Phi_t(\rho) := \Phi(\rho_t)$ and ρ_t is the defined in Lemma 5.8. Also denote by $\pi_{tj} \in \mathbf{P}(E^j)$ the j -th marginal of π_t defined thanks to Theorem 5.1. For any $t > 0$ and any $X^j := (x_1, \dots, x_j) \in E^j$ there holds

$$(5.21) \quad |\nabla_1 \ln \pi_{tj}(X^j)| \leq C_t.$$

PROOF OF LEMMA 5.9. Thanks to Lemma 5.8, we write

$$\begin{aligned} \frac{|\nabla_1 \pi_{tj}(X^j)|}{\pi_{tj}(X^j)} &= \frac{\left| \int \nabla_1 \rho_t(x_1) \rho_t^{\otimes j-1}(x_2, \dots, x_j) \pi(d\rho) \right|}{\pi_{tj}(X^j)} \\ &\leq C_t \frac{\int \rho_t^{\otimes j}(X_j) \pi(d\rho)}{\pi_{tj}(X^j)} = C_t, \end{aligned}$$

which is nothing but (5.21). \square

Lemma 5.10. For any $\pi \in \mathbf{P}(\mathbf{P}(E))$ we define

$$\mathcal{I}'(\pi) := \sup_{j \in \mathbb{N}^*} I(\pi_j).$$

The functional $\mathcal{I}' : \mathbf{P}(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$ is nonnegative, l.s.c. for the weak convergence, satisfies

$$(5.22) \quad \mathcal{I}'(\pi) = \lim_{j \rightarrow \infty} I(\pi_j)$$

and is affine in the same sense as formulated in point (iv) of Lemma 5.6.

PROOF OF LEMMA 5.10. The fact that \mathcal{I}' is nonnegative and l.s.c. is clear and (5.22) comes from the monotony property $I(\pi_{j-1}) \leq I(\pi_j)$, $\forall j \geq 2$ established in Lemma 3.7 (i). It remains only to prove the linearity property of \mathcal{I}' . For the sake of simplicity we only consider the case when $M = 2$ and ω_1 is a ball. The case when ω_1 is a general open set can be handled in a similar way and the case when $M \geq 3$ can be deduced by an iterative argument. For some given $\pi \in \mathbf{P}(\mathbf{P}(E))$ which is not a Dirac mass, $f_1 \in \mathbf{P}(E)$ and $r \in (0, \infty)$ so that

$$\theta := \pi(\mathcal{B}_r) \in (0, 1), \quad \mathcal{B}_r := \mathcal{B}(f_1, r) = \{\rho, W_1(\rho, f_1) < r\},$$

we define

$$F := \frac{1}{\theta} \mathbf{1}_{\mathcal{B}_r} \pi, \quad G := \frac{1}{1-\theta} \mathbf{1}_{\mathcal{B}_r^c} \pi$$

so that

$$F, G \in \mathbf{P}(\mathbf{P}(E)) \quad \text{and} \quad \pi = \theta F + (1-\theta)G,$$

and we have to prove that

$$(5.23) \quad \mathcal{I}'(\pi) = \theta \mathcal{I}'(F) + (1-\theta) \mathcal{I}'(G).$$

We split the proof of that claim in four steps.

Step 1. Approximation and estimation of the affinity defect. As explained for π in the statement of Lemma 5.9, we define F_t and G_t to be the push-forward of the measures F and G by the regularisation operator S_t , and then F_{tj} and G_{tj} are their projections on $\mathbf{P}(E^j)$

$$F_{tj} := \int_{\mathbf{P}(E)} \rho^{\otimes j} F_t(d\rho) = \int_{\mathbf{P}(E)} \rho_t^{\otimes j} F(d\rho), \quad \text{or} \quad \langle F_{tj}, \varphi \rangle = \int_{\mathbf{P}(E)} R_\varphi(\rho) F_t(d\rho),$$

via duality, for any $\varphi \in C_b(E^j)$ where R_φ is the polynomial on $\mathbf{P}(E)$ associated to φ thanks to (2.8). The same holds for G . We also remark that these two above operations (regularisation and projection on E^j) commute if we define the regularisation operators S_t on E^j by the convolution with $\eta_t^{\otimes j}$. It is worth emphasizing that we do not need here, in order to define these objects, that F and G are probability measures, but only that they are Radon measures on $\mathbf{P}(E)$.

For any given $j \in \mathbb{N}$, we define

$$\begin{aligned} A_{tj} &:= \theta I(F_{tj}) + (1-\theta) I(G_{tj}) - I(\theta F_{tj} + (1-\theta) G_{tj}), \\ &= \theta \int \frac{|\nabla F_{tj}|^2}{F_{tj}} + (1-\theta) \int \frac{|\nabla G_{tj}|^2}{G_{tj}} - \int \frac{|(1-\theta)\nabla G_{tj} + \theta\nabla F_{tj}|^2}{(1-\theta)G_{tj} + \theta F_{tj}}. \end{aligned}$$

After reduction to the same denominator, and some simplification, we end up with

$$\begin{aligned} A_{tj} &= \theta(1-\theta) \int \frac{G_{tj} F_{tj}}{(1-\theta)G_{tj} + \theta F_{tj}} \left| \nabla_1 \ln \frac{G_{tj}}{F_{tj}} \right|^2 \\ &\leq 2\theta(1-\theta) \int \frac{G_{tj} F_{tj}}{(1-\theta)G_{tj} + \theta F_{tj}} \left(|\nabla_1 \ln F_{tj}|^2 + |\nabla_1 \ln G_{tj}|^2 \right). \end{aligned}$$

We can estimate the r.h.s. term thanks to Lemma 5.9 by

$$A_{tj} \leq 4\theta(1-\theta)C_t \int \frac{G_{tj}F_{tj}}{(1-\theta)G_{tj} + \theta F_{tj}}.$$

Step 2. Disjunction of the supports. Let us introduce for any $s \in (0, r)$ the two measures on $\mathbf{P}(E)$ (which are not necessarily probability measures)

$$F' := \mathbf{1}_{\mathcal{B}_s} F = \frac{1}{\theta} \mathbf{1}_{\mathcal{B}_s} \pi, \quad F'' := \mathbf{1}_{\mathcal{B}_r \setminus \mathcal{B}_s} F, \quad \text{so that} \quad F' + F'' = F$$

and let us observe that

$$\lim_{s \rightarrow r} \int F''(d\rho) = \lim_{s \rightarrow r} \int \mathbf{1}_{\mathcal{B}_r \setminus \mathcal{B}_s}(\rho) F(d\rho) = 0,$$

by Lebesgue's dominated convergence theorem. For any $t > 0$ and $j \geq 1$ there holds $F'_{tj} + F''_{tj} = F_{tj}$ with $F''_{tj} \geq 0$, so that we may write for any $\varepsilon > 0$

$$\begin{aligned} A_{tj} &\leq 4\theta(1-\theta)C_t \int \frac{G_{tj}F'_{tj}}{(1-\theta)G_{tj} + \theta F'_{tj}} + 4\theta C_t \int F''_{tj}, \\ &\leq 4\theta(1-\theta)C_t \int \frac{G_{tj}F'_{tj}}{(1-\theta)G_{tj} + \theta F'_{tj}} + \varepsilon, \end{aligned}$$

taking s close enough to r , and this independently of j and t because

$$\int_{E^j} F''_{tj} = \int_{\mathbf{P}(E)} F''_t = \int_{\mathbf{P}(E)} F.$$

Step 3. Concentration. We introduce the real numbers $u = \frac{r+s}{2}$ and $\delta = \frac{r-s}{2}$, depending on ε , as well as the set

$$\tilde{\mathcal{B}}_u := \{X^j = (x_1, \dots, x_j), W_1(\mu_{X^j}^j, f_1) < u\} \subset E^j$$

which is nothing but the reciprocal image of the ball $\mathcal{B}_u \subset \mathbf{P}(E)$ by the empirical measure map. Using that

$$\frac{G_{tj}F'_{tj}}{(1-\theta)G_{tj} + \theta F'_{tj}} \leq \frac{1}{\theta} G_{tj} \mathbf{1}_{\tilde{\mathcal{B}}_u} + \frac{1}{1-\theta} F'_{tj} \mathbf{1}_{\tilde{\mathcal{B}}_u^c},$$

we get

$$(5.24) \quad A_{tj} \leq 4C_t \left((1-\theta) \int_{\tilde{\mathcal{B}}_u} G_{tj} + \theta \int_{\tilde{\mathcal{B}}_u^c} F'_{tj} \right) + \varepsilon.$$

If ρ belongs to the support of F' and $X^j \in \tilde{\mathcal{B}}_u^c$, we have thanks to the last estimate in Lemma 5.9

$$\begin{aligned} W_1(\mu_{X^j}^j, \rho_t) &\geq W_1(\mu_{X^j}^j, f_1) - W_1(f_1, \rho) - W_1(\rho, \rho_t) \\ &\geq u - s - C_d t \geq \delta/2, \end{aligned}$$

for any $t \in [0, T(\varepsilon)]$, $T(\varepsilon) > 0$. We first assume that $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ for some $m > 0$, which implies also that $F, G \in \mathbf{P}_m(\mathbf{P}(E))$. Gathering this information with the Chebychev

inequality, estimate (2.30) and estimate (5.20), we conclude that

$$\begin{aligned}
 \int_{\tilde{B}_u^c} F'_{tj} &= \int_{\mathbf{P}(E)} \langle \rho_t^{\otimes j}, \mathbf{1}_{\tilde{B}_u^c} \rangle F'(d\rho) \\
 &\leq \frac{2}{\delta} \int_{\mathbf{P}(E)} \left(\int_{E^j} W_1(\mu_{X^j}^j, \rho_t) \rho_t^{\otimes j}(dX^j) \right) F'(d\rho) \\
 (5.25) \quad &\leq \frac{C}{\delta j^\gamma} \int_{\mathbf{P}(E)} M_m(\rho_t)^{1/m} F'(d\rho) \leq \frac{C}{\delta j^\gamma} \left(\mathcal{M}_m(F) + \varepsilon_{tm} \right)^{1/m},
 \end{aligned}$$

with $\gamma := 1/(d+2+d/m)$. With exactly the same arguments, we prove that for any $\varepsilon > 0$ and any $t \in [0, T(\varepsilon)]$

$$(5.26) \quad \int_{\tilde{B}_u} G_{tj} \leq \frac{2C}{\delta j^\gamma} \mathcal{M}_m(G)^{1/m} + \varepsilon.$$

Gathering (5.24) with (5.25) and (5.26), we get that for any $\varepsilon > 0$, $t \in (0, T(\varepsilon)]$ and $j \geq 1$,

$$A_{tj} \leq \frac{4C_t M_m(\pi)^{1/m}}{\delta j^\gamma} + 3\varepsilon,$$

and then for any $\varepsilon > 0$, $t \in (0, T(\varepsilon)]$

$$(5.27) \quad \limsup_{j \rightarrow \infty} A_{tj} \leq 3\varepsilon.$$

Step 3'. Adaptation for $\pi \notin \mathbf{P}_m(\mathbf{P}(E))$. In the case when $\pi \notin \mathbf{P}_m(\mathbf{P}(E))$ whatever is $m > 0$, we can still prove that (5.27) holds true for any ε and $t \in (0, T(\varepsilon))$ where $T(\varepsilon)$ is small enough. Remark that it cannot be the case if E is a smooth bounded open set, so we have only to deal with the case $E = \mathbb{R}^d$ here.

The idea is the same as in the proof of Lemma 5.6. We choose a function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying (5.15) together with $g(2x) \leq 2g(x)$ for all $x \geq 0$. We argue by using moment with respect to $g(\langle \cdot \rangle)$ rather than to $\langle \cdot \rangle^m$. The property $g(2x) \leq 2g(x)$ ensures that the estimate on the moments in Lemma 5.8 is still true with the moment M_g .

Next, for any $R > 0$, we introduce the mapping from $P_R : \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by

$$P_R(x) := \begin{cases} x & \text{if } |x| \leq R \\ R \frac{x}{|x|} & \text{else} \end{cases}.$$

using the concentration estimate (2.30) for the probability $(\rho \circ P_R^{-1})^{\otimes N} = (P_{R\#}\rho)^{\otimes N}$, it is still possible to deduce that

$$\int_{E^j} W_1(\mu_{X^j}^j, \rho_t) \rho_t^{\otimes j}(dX^j) \leq \frac{C}{\delta j^\gamma} R + \frac{2M_g(\rho_t)}{g(R)}.$$

Summing up with respect to π , choosing R large enough and letting $j \rightarrow +\infty$, we get the claimed inequality for the limsup (with maybe a 4ε in place of the 3ε).

Step 4. Conclusion. The regularization by convolution (or with the heat flow) implies that for any $\alpha \in \mathbf{P}(\mathbf{P}(E))$

$$\mathcal{I}(\alpha_t) = \sup_{j \leq 1} I(\alpha_{tj}) = \sup_{j \leq 1} I(\alpha_{jt}) \leq \sup_{j \leq 1} I(\alpha_j) = \mathcal{I}(\alpha).$$

Moreover, the last point in Lemma 5.8, implies that $\alpha_t \rightarrow \alpha$ with respect to the \mathcal{W}_1 -metric. Thanks to the previous inequality and the l.s.c. property of \mathcal{I}' , we obtain

$$(5.28) \quad \mathcal{I}'(\alpha) = \lim_{t \rightarrow 0} \mathcal{I}'(\alpha_t).$$

Turning back to the definition of A_{tj} , the estimate (5.27) and the above properties, we obtain for any $\varepsilon > 0$, $t \in (0, T(\varepsilon)]$

$$\begin{aligned} \mathcal{I}'(\pi) &\geq \mathcal{I}'(\pi_t) \\ &\geq \theta \mathcal{I}'(F_t) + (1 - \theta) \mathcal{I}'(G_t) - 3\varepsilon. \end{aligned}$$

First passing to the limit $t \rightarrow 0$ and using (5.28) we get

$$\mathcal{I}'(\pi) \geq \theta \mathcal{I}'(F) + (1 - \theta) \mathcal{I}'(G) - 3\varepsilon,$$

for any $\varepsilon > 0$, which concludes the proof of (5.23) since the reverse inequality is just a consequence of the convexity of the functional \mathcal{I}' . \square

PROOF OF THEOREM 5.7. We only have to observe that I_j , \mathcal{I} and \mathcal{I}' fulfil the assumptions of Lemma 5.6. But the assumption (i) is a consequence of Lemma 3.6, the assumption (ii) is a consequence of Lemma 3.5, the assumption (iii) is proved in Lemma 3.6 and assumption (iv) in Lemma 5.10. Then (5.18) and (5.19) are exactly the conclusion of Lemma 5.6 adapted to the Fisher information. \square

Proposition 5.11. *Consider $\pi \in \mathbf{P}(\mathbf{P}(E))$ and (π_j) the associated family of compatible and symmetric probability measures in $\mathbf{P}(E^j)$ defined as in the De Finetti, Hewitt & Savage theorem. For any $p \in [1, +\infty]$, the following equality holds*

$$(5.29) \quad \pi - \text{Suppress} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \} = \sup_{j \in \mathbb{N}} \|\pi_j\|_p^{\frac{1}{j}} = \lim_{j \rightarrow +\infty} \|\pi_j\|_p^{\frac{1}{j}}.$$

It is part of the result that the limit exists. In particular, it implies the equivalence

$$\forall j \in \mathbb{N}, \|\pi_j\|_{L^p(E^j)} \leq C^j \iff \pi - \text{Suppress} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \} \leq C.$$

PROOF OF PROPOSITION 5.11. First remark that there is nothing to prove for $p = 1$ since we are dealing with probability measures. Now, one inequality is a simple consequence of the De Finetti, Hewitt & Savage theorem. In fact, using the definition of π_j , we get

$$\|\pi_j\|_p = \left\| \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \right\|_p \leq \int_{\mathbf{P}(E)} \|\rho^{\otimes j}\|_p \pi(d\rho) = \int_{\mathbf{P}(E)} \|\rho\|_p^j \pi(d\rho),$$

and the last quantity is clearly bounded by M^j , $M := \pi - \text{Suppress} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \}$.

For the reverse inequality, we denote by $q \in (1, +\infty]$ the real conjugate to p . Because $L^q(E) = (L^p(E))'$, the Hahn-Banach separation theorem infers that for any $\lambda < M$ there exists f in the unit ball of $L^q(E)$ so that the set

$$\mathcal{B} := \{ \rho \in \mathbf{P}(E) \text{ s.t. } \int f(x) \rho(dx) \geq \lambda \}$$

is of π -measure positive : $\delta := \int_{\mathcal{B}} \pi(d\rho) > 0$. Now for any $j \in \mathbb{N}$

$$\|\pi_j\|_p \geq \int_{E^j} f^{\otimes j} d\pi_j = \int_{\mathbf{P}(E)} \left(\int_{E^j} f^{\otimes j} \rho^{\otimes j} \right) d\pi(\rho) \geq \delta \lambda^j,$$

which implies the reserve inequality $M \leq \lim_{j \rightarrow +\infty} \|\pi_j\|_p^{\frac{1}{j}}$. \square

5.4. Strong version of De Finetti, Hewitt and Savage theorem and strong convergence in $\mathbf{P}(E^N)$. We begin that section by an HWI inequality valid on $\mathbf{P}(\mathbf{P}(E))$, which is just a "summation" of the usual one and will be very useful in the sequel.

Proposition 5.12. *Assume $E = \mathbb{R}^d$ or more generally that (3.15) holds for $N = 1$. For any $\alpha, \beta \in \mathbf{P}(\mathbf{P}_2(E))$, we have*

$$(5.30) \quad \mathcal{H}(\alpha) \leq \mathcal{H}(\beta) + C_E \sqrt{\mathcal{I}(\alpha)} \mathcal{W}_2(\alpha, \beta).$$

As a consequence, the entropy \mathcal{H} is continuous on bounded sets relatively to \mathcal{I} . In more precise words, if (π_n) is a bounded sequence of $\mathbf{P}_m(\mathbf{P}(E))$, $m > 0$, such that

$$\pi_n \rightharpoonup \pi \text{ weakly in } \mathbf{P}(\mathbf{P}(E)) \quad \text{and} \quad \mathcal{I}(\pi_n) \leq C,$$

then $\mathcal{H}(\pi_n) \rightarrow \mathcal{H}(\pi)$.

PROOF OF PROPOSITION 5.12. A first way in order to prove (5.30) is just to pass in the limit in the HWI inequality (3.15) for α_N and β_N and use the inequality stated in lemma 2.7 for the quadratic cost, and the result of the previous section about level 3 entropy and Fisher information 5.9 et 5.18.

Another possibility is to sum up the HWI inequality (3.16) for $\rho \in \mathbf{P}(E)$. Choosing an optimal transference plan Π for \mathcal{W}_2 between α and β , we have

$$\int_{\mathbf{P}(E)} H(\rho) \Pi(d\rho, d\eta) \leq \int_{\mathbf{P}(E)} H(\eta) \Pi(d\rho, d\eta) + \int_{\mathbf{P}(E)} \sqrt{\mathcal{I}(\rho)} \mathcal{W}_2(\rho, \eta) \Pi(d\rho, d\eta),$$

so that

$$H(\alpha) \leq H(\beta) + \left(\int_{\mathbf{P}(E)} \mathcal{I}(\rho) \Pi(d\rho, d\eta) \right)^{\frac{1}{2}} \left(\int_{\mathbf{P}(E)} \mathcal{W}_2(\rho, \eta)^2 \Pi(d\rho, d\eta) \right)^{\frac{1}{2}},$$

thanks to Cauchy-Schwarz inequality. It leads to the desired inequality.

The second point is obtained by two applications of the previous inequality, leading to

$$|\mathcal{H}(\pi_n) - \mathcal{H}(\pi)| \leq \left(\sqrt{\mathcal{I}(\pi)} + \sqrt{\mathcal{I}(\pi_n)} \right) \mathcal{W}_2(\pi_n, \pi),$$

and then using the l.s.c. property of the level 3 Fisher information in order to prove that $\mathcal{I}(\pi) < \infty$. We conclude by remarking that the RHS converges to 0 as n tends to ∞ . \square

The results of the preceding section and the HWI inequality make possible to compare different senses of convergence for sequences of $\mathbf{P}(E^N)$, $N \rightarrow \infty$, without any assumption of chaos.

Theorem 5.13. *Assume $E = \mathbb{R}^d$ or $E \subset \mathbb{R}^d$ is a bounded connected open subset with smooth boundary and that (3.15) holds. Consider (F^N) a sequence of $\mathbf{P}_{\text{sym}}(E^N)$ and $\pi \in \mathbf{P}(\mathbf{P}(E))$ such that $F^N \rightharpoonup \pi$ weakly in $\mathbf{P}_k(E^j)_{\forall j}$, $k > 2$.*

(1) *In the list of assertions below, each assertion implies the one which follows:*

- (i) $\mathcal{I}(F^N) \rightarrow \mathcal{I}(\pi)$, $\mathcal{I}(\pi) < \infty$;
- (ii) $\mathcal{I}(F^N)$ is bounded;
- (iii) $\mathcal{H}(F^N) \rightarrow \mathcal{H}(\pi)$, $\mathcal{H}(\pi) < \infty$.

(2) *More precisely, the following version of the implication (ii) \Rightarrow (iii) holds. There exists a numerical constant C such that for any $k > 2$ and $K > 0$, and for any any sequence (F^N) of $\mathbf{P}_{\text{sym}}(E^N)$ satisfying*

$$\forall N \quad M_k(F_1^N) \leq K^k, \quad \mathcal{I}(F^N) \leq K^2,$$

there holds

$$(5.31) \quad \forall N \geq 4^{2d} \quad |H(F^N) - \mathcal{H}(\pi)| \leq K W_2(F^N, \pi_N) + CK^{d'} \frac{\ln(KN)}{N^\gamma},$$

with $\gamma := \frac{k-2}{k(1+2d)+4d-2}$ and as usual $d' = \max(2, d)$.

(3) In particular, for any sequence (π_j) of symmetric and compatible probability measures of $\mathbf{P}(E^j)$ satisfying

$$M_k(\pi_1) \leq K^k, \quad \forall j \geq 1 \quad I(\pi_j) \leq K^2,$$

there holds

$$(5.32) \quad \forall j \geq 4^{2d}, \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq CK^{d'} \frac{\ln(Kj)}{j^\gamma}$$

for the same value of γ . In other words, (5.32) gives a rate of convergence for the limit (5.9).

The fact that the constant C does not depend on k is interesting when the space E is compact or the measures F^N have strong integrability properties, for instance an exponential moment. It allows to choose large k and get almost the largest exponent γ possible. Precise versions of the point (iii) are stated (without proofs) in the corollary below.

Corollary 5.14. (i) In the case where E is compact, we denote $K := \max(\text{diam}(E), \sqrt{\mathcal{I}(\pi)})$. Then there holds for all $j \geq 4^{2d}$

$$(5.33) \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq CK^{d'} \frac{\ln(Kj)}{j^\gamma} \quad \text{with } \gamma = \frac{1}{2d+1}$$

(ii) If $M_{\beta, \lambda}(\pi_1) := \int_E e^{\lambda|x|^\beta} \pi_1(dx) < +\infty$ for some $\lambda > 0$ and $\mathcal{I}(\pi) < +\infty$, there exists a constant $C(d, \beta, \lambda, \mathcal{I}(\pi))$ such that for j large enough ($\geq C' \ln M_{\beta, \lambda}(\pi_1)$)

$$(5.34) \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq C \frac{[\ln j]^{1+d'/\beta}}{j^\gamma} \quad \text{with } \gamma = \frac{1}{2d+1}.$$

PROOF OF THEOREM 5.13. We split the proof into four steps.

Step 1. i) implies ii) is clear. For ii) implies iii), we use the HWI inequality (3.15) and we write

$$\begin{aligned} |H(F^N) - \mathcal{H}(\pi)| &= |H(F^N) - H(\pi_N) + H(\pi_N) - \mathcal{H}(\pi)| \\ &\leq C_E \left(\sqrt{I(F^N)} + \sqrt{I(\pi_N)} \right) W_2(F^N, \pi_N) + |H(\pi_N) - \mathcal{H}(\pi)|. \end{aligned}$$

We know from (5.9) that $\mathcal{H}(\pi) = \lim H(\pi_N)$ and from (5.18) and (5.19) that $I(\pi_N) \leq \mathcal{I}(\pi) \leq \liminf I(F^N) \leq K$, from which we conclude that there exist a sequence $\varepsilon_\pi(N) \rightarrow 0$ such that

$$|H(F^N) - \mathcal{H}(\pi)| \leq 2C_E K W_2(F^N, \pi_N) + \varepsilon(N).$$

We now aim to estimate $\varepsilon(N)$ more explicitly as claimed in point **(3)**. Then **(2)** will be a direct consequence of **(3)** and the above estimate.

From now on, we only consider the case $E = \mathbb{R}^d$ since the general case is similar (and the case when E is compact is even simpler).

Step 2. From [12, Theorem A.1] we know that for any $R, \delta > 0$ we may cover $\mathbf{P}(B_R)$ by $\mathcal{N}(R, \delta/2)$ balls of radius $\delta/2$ in W_1 distance (which is less accurate than the one considered in the above quoted result) with

$$\mathcal{N}(R, \delta) \leq \left(\frac{C'_1 R}{\delta} \right)^{C'_2 (R/\delta)^d},$$

where the constant C'_1 and C'_2 are numerical. Let us fix $a \geq 1$ and recall that we define $\mathcal{BP}_{k,a}(E) := \{\rho \in \mathbf{P}(E) \text{ s.t. } M_k(\rho) \leq a\}$. Next, for any $\rho \in \mathcal{BP}_{k,a}(E)$, we define $\rho_R \in \mathbf{P}(B_R)$ by $\rho_R = \rho(B_R)^{-1} \rho \mathbf{1}_{B_R}$ for R large enough (so that it defines a probability measure), and we observe that for any $f \in \mathbf{P}(E)$ we have

$$W_1(\rho, f) \leq W_1(\rho_R, f) + W_1(\rho_R, \rho),$$

and that for any R such that $R^k > 2a$

$$\begin{aligned} W_1(\rho_R, \rho) &\leq \|\rho_R - \rho\|_{TV} \leq \left| 1 - \frac{1}{\rho(B_R)} \right| + \rho(B_R^c) \\ &\leq \left(1 + \frac{1}{\rho(B_R)} \right) \rho(B_R^c) \leq 3 \frac{a}{R^k}, \end{aligned}$$

since then $\rho(B_R) \geq 1 - \frac{a}{R^k} \geq \frac{1}{2}$.

As a consequence, for any $\delta \leq 1$ and $a \geq 1$, choosing R such that $3a/R^k = \delta/2$ in the two preceding estimates, we may cover $\mathcal{BP}_{k,a}(E)$ by $\mathcal{N}_a(\delta) = \mathcal{N}(R, \delta/2)$ balls of radius δ in W_1 distance, with

$$\frac{1}{\delta} \leq \mathcal{N}_a(\delta) \leq \left(C_1 a^{\frac{1}{k}} \delta^{-1-\frac{1}{k}} \right)^{C_2 a^{\frac{d}{k}} \delta^{-d-\frac{d}{k}}}.$$

The above lower bound on $\mathcal{N}_a(\delta)$ is straightforwardly obtained by considering balls centered on Dirac masses distributed on a line. In the sequel, we shall often use the shortcut $\mathcal{N} = \mathcal{N}_a(\delta)$. Let us then introduce a covering family $\omega_i^\delta \subset \mathcal{BP}_{k,a}(E)$, $1 \leq i \leq \mathcal{N}_a(\delta)$, such that

$$\sup_{\rho, \eta \in \omega_i^\delta} W_1(\rho, \eta) \leq 2\delta, \quad \omega_i^\delta \cap \omega_j^\delta = \emptyset \text{ if } i \neq j, \quad \mathcal{BP}_{k,a}(E) = \bigcup_{i=1}^{\mathcal{N}_a(\delta)} \omega_i^\delta,$$

as well as the masses and centers of mass

$$\alpha_i^\delta := \int_{\omega_i^\delta} \pi, \quad f_i^\delta := \frac{1}{\alpha_i^\delta} \int_{\omega_i^\delta} \rho \pi(d\rho).$$

We also denote $\omega_0^\delta := [\mathcal{BP}_{k,a}(E)]^c$ and $\alpha_0^\delta := \int_{\omega_0^\delta} \pi$, so that $\sum_{i=0}^{\mathcal{N}} \alpha_i^\delta = 1$. Denoting $\mathcal{Z} := \{i = 1, \dots, \mathcal{N}_a(\delta); \alpha_i^\delta \geq \mathcal{N}_a(\delta)^{-2}\}$, we finally define

$$\pi^\delta := \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta \delta_{f_i^\delta}, \quad \text{with } \beta_i^\delta := \frac{\alpha_i^\delta}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \text{ if } i \in \mathcal{Z} \text{ and } \beta_i^\delta := 0 \text{ if } i \notin \mathcal{Z}.$$

Remark that by our moment assumption

$$\alpha_0^\delta \leq \int_{(\omega_0^\delta)^c} \pi(d\rho) \leq \int_{\mathbf{P}(E)} \frac{M_k(\rho)}{a} \pi(d\rho) = \frac{M_k(\pi_1)}{a}.$$

Since $\sum_{i \notin \mathcal{Z}, i \geq 1} \alpha_i^\delta \leq \mathcal{N}^{-1} \leq \delta$, we necessarily have $\mathcal{Z} \neq \emptyset$ if $\delta + \frac{M_k(\pi_1)}{a} \leq \frac{1}{2} < 1$, an assumption that we will make in the sequel. We fix now the value of a to be so that

$$\delta = \frac{M_k(\pi_1)}{a}.$$

As we shall see that will lead to the optimal inequality. With that particular choice, the condition above simply writes $\delta \leq \frac{1}{4}$, and the upper bound on \mathcal{N} may be rewritten

$$(5.35) \quad \mathcal{N}(\delta) := \mathcal{N}_a(\delta) \leq \left(C_1 K \delta^{-1-\frac{2}{k}} \right)^{C_2 K^d \delta^{-d(1+\frac{2}{k})}}.$$

In that case, we have

$$(5.36) \quad \sum_{j \in \mathcal{Z}, j \geq 0} \alpha_j^\delta \leq 2\delta, \quad 1 \geq \sum_{j \in \mathcal{Z}} \alpha_j^\delta \geq 1 - 2\delta \geq \frac{1}{2}.$$

Now, by convexity of the Fisher information

$$I(f_i^\delta) \leq \frac{1}{\alpha_i^\delta} \int_{\omega_i^\delta} I(\rho) \pi(d\rho),$$

which in turns implies that

$$\mathcal{I}(\pi^\delta) = \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta I(f_i^\delta) \leq \frac{1}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \sum_{i \in \mathcal{Z}} \int_{\omega_i^\delta} I(\rho) \pi(d\rho) \leq 2\mathcal{I}(\pi).$$

Similarly, for the moment of order k :

$$M_k(\pi_1^\delta) = \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta M_k(f_i^\delta) \leq \frac{1}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \sum_{i \in \mathcal{Z}} \int_{\omega_i^\delta} M_k(\rho) \pi(d\rho) \leq 2 M_k(\pi_1).$$

In order to prove (5.32), we introduce the splitting

$$(5.37) \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq |H(\pi_j) - H(\pi_j^\delta)| + |H(\pi_j^\delta) - \mathcal{H}(\pi^\delta)| + |\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)|,$$

where we have written $\pi_j^\delta := (\pi^\delta)_j$. We now estimate each term separately.

Step 3. On the one hand, defining $T^\delta : \mathbf{P}(E) \rightarrow \{f_0^\delta, \dots, f_{\mathcal{N}}^\delta\}$, $T^\delta(\rho) = f_i^\delta$ if $\rho \in \omega_i^\delta$, $T^\delta(\rho) = f_0^\delta = \delta_0$ if $\rho \in \omega_0^\delta$ and $\beta_0^\delta := 0$, we compute

$$\begin{aligned} \mathcal{W}_1(\pi, \pi^\delta) &\leq \mathcal{W}_1\left(\pi, \sum_{i=0}^{\mathcal{N}} \alpha_i^\delta \delta_{f_i^\delta}\right) + \mathcal{W}_1\left(\sum_{i=0}^{\mathcal{N}} \alpha_i^\delta \delta_{f_i^\delta}, \sum_{i=1}^{\mathcal{N}} \beta_i^\delta \delta_{f_i^\delta}\right) \\ &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho, \eta) (Id \otimes T^\delta) \# \pi + \left\| \sum_{i=0}^{\mathcal{N}} (\alpha_i^\delta - \beta_i^\delta) \delta_{f_i^\delta} \right\|_{TV} \\ &\leq \int_{\mathbf{P}(E)} W_1(\rho, T^\delta(\rho)) \pi(d\rho) + \sum_{i=1}^{\mathcal{N}} |\alpha_i^\delta - \beta_i^\delta| + |\alpha_0^\delta| \\ &\leq \delta + \frac{M_1(\pi)}{a} + 6\delta \leq 8\delta, \end{aligned}$$

where we have used several times estimation (5.36), in particular in order to get the inequality

$$\sum_{i=1}^{\mathcal{N}} |\alpha_i^\delta - \beta_i^\delta| + \alpha_0^\delta = \left(1 - \frac{1}{\sum_{i \in \mathcal{Z}} \alpha_i^\delta} \right) \sum_{i \in \mathcal{Z}} \alpha_i^\delta + \sum_{i \notin \mathcal{Z}} \alpha_i^\delta \leq 3 \sum_{i \notin \mathcal{Z}} \alpha_i^\delta.$$

Using lemma 2.3 and the bound on $M_k(\pi_1^\delta)$, we obtain a bound on $\mathcal{W}_2(\pi^\delta, \pi)$ as follows (we recall that the constant C that appears is numerical : $C = 2^{3/2}$)

$$\mathcal{W}_2(\pi^\delta, \pi) \leq C M_k(\pi_1)^{1/k} \mathcal{W}_1(\pi^\delta, \pi)^{1/2-1/k} \leq 4 C K \delta^{1/2-1/k}.$$

Now, we use the HWI inequality on $\mathbf{P}(E)$ stated in Proposition 5.12 and we bound the first term in (5.37) by

$$|\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)| \leq \left[\sqrt{\mathcal{I}(\pi^\delta)} + \sqrt{\mathcal{I}(\pi)} \right] \mathcal{W}_2(\pi^\delta, \pi) \leq 2 K \mathcal{W}_2(\pi^\delta, \pi),$$

and the third term in (5.37) very similarly

$$\begin{aligned} |H(\pi_j^\delta) - \mathcal{H}(\pi_j)| &\leq \left[\sqrt{I(\pi_j^\delta)} + \sqrt{I(\pi_j)} \right] W_2(\pi_j^\delta, \pi_j) \\ &\leq \left[\sqrt{\mathcal{I}(\pi^\delta)} + \sqrt{\mathcal{I}(\pi)} \right] \mathcal{W}_2(\pi^\delta, \pi) \leq 2 K \mathcal{W}_2(\pi^\delta, \pi), \end{aligned}$$

where we have used the properties (5.18) of the level 3 Fisher information and Lemma 2.7 in order to bound W_2 by \mathcal{W}_2 . All together, we have proved

$$(5.38) \quad |\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)| + |H(\pi_j^\delta) - \mathcal{H}(\pi_j)| \leq C K^2 \delta^{1/2-1/k},$$

for some numerical constant $C \leq 2^6$.

Step 4. We estimate the second term in (5.37). Using that $\pi_j^\delta = \beta_1^\delta (f_1^\delta)^{\otimes j} + \dots + \beta_{\mathcal{N}}^\delta (f_{\mathcal{N}}^\delta)^{\otimes j}$, we write

$$\begin{aligned} H(\pi_j^\delta) &= \frac{1}{j} \int_{E^j} \pi_j^\delta \log \pi_j^\delta \\ &= \sum_{i=1}^{\mathcal{N}} \beta_i^\delta H(f_i^\delta) + \frac{1}{j} \int_{E^j} \pi_j^\delta \Lambda \left(\frac{\beta_1^\delta (f_1^\delta)^{\otimes j}}{\pi_j^\delta}, \dots, \frac{\beta_{\mathcal{N}}^\delta (f_{\mathcal{N}}^\delta)^{\otimes j}}{\pi_j^\delta} \right), \end{aligned}$$

with $\Lambda : \{U = (u_i) \in \mathbb{R}_+^{\mathcal{N}}, \sum_i u_i = 1\} \rightarrow \mathbb{R}$ defined by

$$\Lambda(U) := u_1 \log \left(\frac{\beta_1^\delta}{u_1} \right) + \dots + u_{\mathcal{N}} \log \left(\frac{\beta_{\mathcal{N}}^\delta}{u_{\mathcal{N}}} \right).$$

Observing that Λ is in fact (the opposite of) a discrete relative entropy, we have for any $U \in \mathbb{R}_+^{\mathcal{N}}$ with $\sum_i u_i = 1$

$$-\log(\mathcal{N}^2) \leq \log(\min \beta_i^\delta) \leq \Lambda(U) \leq 0,$$

we deduce

$$|H(\pi_j^\delta) - \mathcal{H}(\pi_j)| \leq \frac{2}{j} \log \mathcal{N}_a(\delta).$$

Step 5. All in all, observing that thanks to (5.35)

$$\log \mathcal{N}(\delta) \leq C K^d \delta^{-d(1+\frac{2}{k})} [1 + \ln K - \ln \delta],$$

we have

$$C^{-1} |H(\pi_j) - \mathcal{H}(\pi)| \leq K^2 \delta^{1/2-1/k} + \frac{1}{j} \frac{K^d}{\delta^{d(1+\frac{2}{k})}} [1 + \ln(K \delta^{-1})].$$

We can now (almost) optimize by choosing $\delta = j^{-r}$, with $r^{-1} := \frac{1}{2} - \frac{1}{k} + d(1 + \frac{2}{k})$ we obtain

$$C^{-1} |H(\pi_j) - \mathcal{H}(\pi)| \leq K^{\max(2,d)} \frac{\ln(Kj)}{j^\gamma}$$

for the integers $j \geq 4^{1/r}$ so that the condition $\delta \leq \frac{1}{4}$ is fulfilled (in order to ensure that $\mathcal{Z} \neq \emptyset$). But it can be checked that for $k \in [2, +\infty)$, $d \leq \frac{1}{r} \leq 2d$. So that the previous condition on j is fulfilled for $j \geq 4^{2d}$. \square

REFERENCES

- [1] AJTAI, M., KOMLÓS, J., AND TUSNÁDY, G. On optimal matchings. *Combinatorica* 4, 4 (1984), 259–264.
- [2] ARKERYD, L., CAPRINO, S., AND IANIRO, N. The homogeneous Boltzmann hierarchy and statistical solutions to the homogeneous Boltzmann equation. *J. Statist. Phys.* 63, 1-2 (1991), 345–361.
- [3] ARTSTEIN, S., BALL, K. M., BARTHE, F., AND NAOR, A. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields* 129, 3 (2004), 381–390.
- [4] BARTHE, F., AND BORDENAVE, C. Combinatorial optimization over two random point sets. arXiv:1103.2734v2.
- [5] BARTHE, F., CORDERO-ERAUSQUIN, D., AND MAUREY, B. Entropy of spherical marginals and related inequalities. *J. Math. Pures Appl. (9)* 86, 2 (2006), 89–99.
- [6] BEN AROUS, G., AND ZEITOUNI, O. Increasing propagation of chaos for mean field models. *Ann. Inst. H. Poincaré Probab. Statist.* 35, 1 (1999), 85–102.
- [7] BERRY, A. C. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* 49 (1941), 122–136.
- [8] BOBKOV, S. G., CHISTYAKOV, G., AND GÖTZE, F. Berry-Esseen bounds in the entropic central limit theorem. arXiv:1104.3994v1.
- [9] BOBKOV, S. G., GENTIL, I., AND LEDOUX, M. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)* 80, 7 (2001), 669–696.
- [10] BOISSARD, E., AND LE GOUIC, T. On the mean speed of convergence of empirical and occupation measures in wasserstein distance. arXiv:1105.5263v1.
- [11] BOLLEY, F., GUILLIN, A., AND MALRIEU, F. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.* 44, 5 (2010), 867–884.
- [12] BOLLEY, F., GUILLIN, A., AND VILLANI, C. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* 137, 3-4 (2007), 541–593.
- [13] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. *Comm. Math. Phys.* 143, 3 (1992), 501–525.
- [14] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. II. *Comm. Math. Phys.* 174, 2 (1995), 229–260.
- [15] CANTELLI, F. P. Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari* 4 (1933), 421–424.
- [16] CARLEN, E. A. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.* 101, 1 (1991), 194–211.
- [17] CARLEN, E. A., CARVALHO, M. C., LE ROUX, J., LOSS, M., AND VILLANI, C. Entropy and chaos in the Kac model. *Kinet. Relat. Models* 3, 1 (2010), 85–122.
- [18] CARLEN, E. A., LIEB, E. H., AND LOSS, M. A sharp analog of Young’s inequality on S^N and related entropy inequalities. *J. Geom. Anal.* 14, 3 (2004), 487–520.
- [19] CARRAPATOSO, K. Quantitative and qualitative Kac’s chaos on the Boltzmann’s sphere. arXiv:1205.1241.
- [20] CARRAPATOSO, K. Propagation of chaos for the spatially homogeneous Landau equation for Maxwellian molecules. arXiv:1212.3724 sphere. arXiv:1205.1241.
- [21] CARRILLO, J., AND TOSCANI, G. Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Rivista Matematica di Parma* 6 (2007), 75–198.
- [22] CHOQUET, G., AND MEYER, P.-A. Existence et unicité des représentations intégrales dans les convexes compacts quelconques. *Ann. Inst. Fourier (Grenoble)* 13 (1963), 139–154.
- [23] CORDERO-ERAUSQUIN, D. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* 161, 3 (2002), 257–269.

- [24] DE FINETTI, B. La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* 7, 1 (1937), 1–68.
- [25] DIACONIS, P., AND FREEDMAN, D. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.* 23, 2, suppl. (1987), 397–423.
- [26] DOBRIĆ, V., AND YUKICH, J. E. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8, 1 (1995), 97–118.
- [27] DOBRUŠIN, R. L. Vlasov equations. *Funktsional. Anal. i Prilozhen.* 13, 2 (1979), 48–58, 96.
- [28] DUDLEY, R. M. The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist* 40 (1968), 40–50.
- [29] ESSEEN, C.-G. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* 28A, 9 (1942), 19.
- [30] FELLER, W. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.
- [31] FONSECA, I., AND PARRY, G. Equilibrium configurations of defective crystals. *Arch. Rational Mech. Anal.* 120, 3 (1992), 245–283.
- [32] FOURNIER, N., HAURAY, M., AND MISCHLER, S. Propagation of chaos for the 2d viscous vortex model. hal-00762286.
- [33] GLIVENKO, V. Sulla determinazione empirica della legge di probabilita. *Giorn. Ist. Ital. Attuari* 4 (1933), 92–99.
- [34] GOUDON, T., JUNCA, S., AND TOSCANI, G. Fourier-based distances and Berry-Esseen like inequalities for smooth densities. *Monatsh. Math.* 135, 2 (2002), 115–136.
- [35] GRAHAM, C., AND MÉLÉARD, S. Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *The Annals of Probability* 25 (1997), 115–132.
- [36] GRÜNBAUM, F. A. Propagation of chaos for the Boltzmann equation. *Arch. Rational Mech. Anal.* 42 (1971), 323–345.
- [37] GUO, M. Z., PAPANICOLAOU, G. C., AND VARADHAN, S. R. S. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys.* 118, 1 (1988), 31–59.
- [38] HAURAY, M., AND JABIN, P.-E. Particles approximations of Vlasov equations with singular forces : Part. 2. hal-00609453.
- [39] HAURAY, M., AND JABIN, P.-E. N -particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.* 183, 3 (2007), 489–524.
- [40] HEWITT, E., AND SAVAGE, L. J. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* 80 (1955), 470–501.
- [41] KAC, M. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III* (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.
- [42] KOSYGINA, E. The behavior of the specific entropy in the hydrodynamic scaling limit. *Ann. Probab.* 29, 3 (2001), 1086–1110.
- [43] LANFORD, III, O. E. Time evolution of large classical systems. In *Dynamical systems, theory and applications (Recontres, Battelle Res. Inst., Seattle, Wash., 1974)*. Springer, Berlin, 1975, pp. 1–111. Lecture Notes in Phys., Vol. 38.
- [44] LIONS, P.-L. Théorie des jeux de champ moyen et applications (mean field games). In *Cours du Collège de France*. http://www.college-de-france.fr/default/EN/all/equ_der/audio_video.jsp, 2007–2009.
- [45] LIONS, P.-L., AND TOSCANI, G. A strengthened central limit theorem for smooth densities. *J. Funct. Anal.* 129, 1 (1995), 148–167.
- [46] LOTT, J., AND VILLANI, C. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)* 169, 3 (2009), 903–991.
- [47] MALRIEU, F. Logarithmic Sobolev inequalities for some nonlinear PDE’s. *Stochastic Process. Appl.* 95, 1 (2001), 109–132.
- [48] MCKEAN, JR., H. P. Speed of approach to equilibrium for Kac’s caricature of a Maxwellian gas. *Arch. Rational Mech. Anal.* 21 (1966), 343–367.
- [49] MCKEAN, JR., H. P. An exponential formula for solving Boltzmann’s equation for a Maxwellian gas. *J. Combinatorial Theory* 2 (1967), 358–382.
- [50] MCKEAN, H. P. The central limit theorem for Carleman’s equation. *Israel J. Math.* 21, 1 (1975), 54–92.
- [51] MEHLER, F. G. Ueber die entwicklung einer function von beliebig vielen variablen nach laplaschen functionen höherer ordnungn. *Crelle’s Journal* 66 (1866), 161–176.

- [52] MÉLÉARD, S. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, vol. 1627 of *Lecture Notes in Math.* Springer, Berlin, 1996, pp. 42–95.
- [53] MESSER, J., AND SPOHN, H. Statistical mechanics of the isothermal Lane-Emden equation. *J. Statist. Phys.* 29, 3 (1982), 561–578.
- [54] MISCHLER, S. Le programme de Kac sur les limites de champ moyen. Séminaire EDP-X, Décembre 2010.
- [55] MISCHLER, S., AND MOUHOT, C. Kac’s program in kinetic theory. *To appear in Inventiones Mathematicae* <http://dx.doi.org/10.1007/s00222-012-0422-3>
- [56] MISCHLER, S., AND MOUHOT, C. On chaos properties of a Boltzmann-Kac system for inelastic collisions. Work in progress.
- [57] MISCHLER, S., MOUHOT, C., AND WENNBURG, B. A new approach to quantitative chaos propagation for drift, diffusion and jump processes. hal-00559132.
- [58] OLLA, S., VARADHAN, S. R. S., AND YAU, H.-T. Hydrodynamical limit for a Hamiltonian system with weak noise. *Comm. Math. Phys.* 155, 3 (1993), 523–560.
- [59] OSADA, H. Propagation of chaos for the two-dimensional Navier-Stokes equation. *Proc. Japan Acad. Ser. A Math. Sci.* 62, 1 (1986), 8–11.
- [60] OTTO, F. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* 26, 1-2 (2001), 101–174.
- [61] OTTO, F., AND VILLANI, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173, 2 (2000), 361–400.
- [62] RACHEV, S. T., AND RÜSCHENDORF, L. *Mass transportation problems. Vol. II. Probability and its Applications* (New York). Springer-Verlag, New York, 1998. Applications.
- [63] RIO, E. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* 45, 3 (2009), 802–817.
- [64] ROBINSON, D. W., AND RUELLE, D. Mean entropy of states in classical statistical mechanics. *Comm. Math. Phys.* 5 (1967), 288–300.
- [65] SZNITMAN, A.-S. Équations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete* 66, 4 (1984), 559–592.
- [66] SZNITMAN, A.-S. A propagation of chaos result for Burgers’ equation. In *Hydrodynamic behavior and interacting particle systems (Minneapolis, Minn., 1986)*, vol. 9 of *IMA Vol. Math. Appl.* Springer, New York, 1987, pp. 181–188.
- [67] SZNITMAN, A.-S. Topics in propagation of chaos. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math.* Springer, Berlin, 1991, pp. 165–251.
- [68] TALAGRAND, M. Matching random samples in many dimensions. *Ann. Appl. Probab.* 2, 4 (1992), 846–856.
- [69] TANAKA, H. Some probabilistic problems in the spatially homogeneous Boltzmann equation. In *Theory and application of random fields (Bangalore, 1982)*, vol. 49 of *Lecture Notes in Control and Inform. Sci.* Springer, Berlin, 1983, pp. 258–267.
- [70] TOSCANI, G. New a priori estimates for the spatially homogeneous Boltzmann equation. *Contin. Mech. Thermodyn.* 4, 2 (1992), 81–93.
- [71] VARADARAJAN, V. S. On the convergence of sample probability distributions. *Sankhyā. The Indian Journal of Statistics* 19 (1958), 23–26.
- [72] VILLANI, C. *Topics in Optimal Transportation*, vol. 58 of *Graduate Studies in Mathematics series*. American Mathematical Society, 2003.
- [73] VILLANI, C. *Optimal transport, old and new*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.
- [74] VILLANI, C. Fisher information estimates for Boltzmann’s collision operator. *J. Math. Pures Appl.* 77 (1998), no. 8, 821–837.
- [75] VILLANI, C. Decrease of the Fisher information for solutions of the spatially homogeneous Landau equation with Maxwellian molecules. *Math. Models Methods Appl. Sci.* 10 (2000), no. 2, 153–161.
- [76] YAU, H.-T. Relative entropy and hydrodynamics of Ginzburg-Landau models. *Lett. Math. Phys.* 22, 1 (1991), 63–80.

MAXIME HAURAY

UNIVERSITÉ D’AIX-MARSEILLE

LATP, UMR CNRS 7353
CENTRE DE MATHÉMATIQUES ET INFORMATIQUE (CMI)
39, RUE F. JOLIOT CURIE 13453 MARSEILLE CEDEX 13
FRANCE

E-MAIL: `maxime.hauray@univ-amu.fr`

STÉPHANE MISCHLER

UNIVERSITÉ PARIS-DAUPHINE & IUF
CEREMADE, UMR CNRS 7534
PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY 75775 PARIS CEDEX 16
FRANCE

E-MAIL: `mischler@ceremade.dauphine.fr`